

# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 136995**

**TO: Ginny Portner**  
**Location: REM-3B02&3C18**  
**Art Unit: 1645**  
**Sunday, November 14, 2004**

**Case Serial Number: 09/545199**

**From: Toby Port**  
**Location: Biotech-Chem Library**  
**REM-1A59**  
**Phone: 571-272-2523**

**toby.port@uspto.gov**

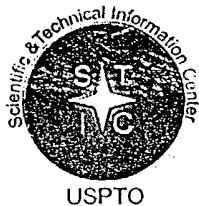
### **Search Notes**

Examiner Portner,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Toby Port



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact*:

Mary Hale, Information Branch Supervisor  
571-272-2507 Remsen E01 D86

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



136995

STIC-Biotech/ChemLib

From: Portner, Ginny  
Sent: Wednesday, November 03, 2004 2:23 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/545,199

Please search SEQ ID NO 4, and back translate the amino acid to the corresponding nucleotide sequence. The claims recite 70% identity to a nucleotide sequence that encodes SEQ Id NO 4. Thanks

Ginny Portner  
Remsen Building  
Art Unit 1645  
Room E03, B02  
(571) 272-0862

CRFE

\*\*\*\*\*  
STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA Sequence: # \_\_\_\_\_  
AA Sequence : # \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2004, 09:43:28 ; Search time 155 Seconds  
(without alignments)  
668.857 Million cell updates/sec

Title: US-09-545-199F-4

Perfect score: 1429

Sequence: 1 MAGAKERTRKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseqp\_23Sep04:\*

2: Geneseqp1980s:\*

3: Geneseqp1990s:\*

4: Geneseqp2000s:\*

5: Geneseqp2001s:\*

6: Geneseqp2002s:\*

7: Geneseqp2003as:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

result No.	Score	Query Match %	Length	ID	Description
1	1429	100.0	289	3	AAB44522
2	1429	100.0	289	5	ABP54474
3	1132.5	79.3	288	3	AAB44579
4	1132.5	79.3	288	5	ABP54531
5	1077.5	75.4	288	5	ABP54547
6	1004	70.3	288	6	ABM70561
7	1004	70.3	291	7	ADF07164
8	991	69.3	287	7	ABG98289
9	978	68.4	297	7	ABO64099
10	926.5	64.8	291	7	ADP70334
11	894	62.6	292	6	ADP33307
12	885	61.9	309	8	ADL05162
13	774	54.2	269	4	ABG24481
14	760	53.2	291	6	ABP77380
15	642.5	45.0	924	4	ABG13826
16	616	43.1	794	4	ABG25667
17	530.5	37.1	304	7	ADH88017
18	524.5	36.7	289	2	AAW56792
19	524.5	36.7	556	4	ABG18379
20	521.5	36.5	288	4	ABG55114
21	518.5	36.3	293	5	ABP27578
22	510.5	35.7	300	7	ADC97657
23	508.5	35.6	291	5	ABP27579
24	504.5	35.3	289	2	AAW56795
25	501.5	35.1	292	6	ABU01963

26	501.5	35.1	292	6	ABP81612
27	501.5	35.1	292	8	ADK46870
28	493	34.5	295	5	ABP39397
29	489	34.2	290	5	ABBA7870
30	486	34.0	288	6	ABM71838
31	464	32.5	272	4	AAAG82359
32	461.5	32.3	325	4	AAAB79630
33	461.5	32.3	325	4	AAAG91087
34	460.5	32.2	293	2	AAW56798
35	454	31.8	365	3	AAAG04932
36	454	31.8	373	3	AAAG04931
37	454	31.8	373	8	ADN74311
38	448	31.4	357	7	ABM73921
39	447.5	31.3	327	5	ABM80865
40	440	30.8	308	7	ADB74399
41	416.5	29.1	315	4	AAU61796
42	416.5	29.1	315	6	ABM58315
43	391	27.4	301	3	AAAG04933
44	355.5	24.9	307	5	ABP65479
45	350.5	24.5	323	3	AAAG08747

## ALIGNMENTS

RESULT 1

AAB44522  
ID AAB44522 standard; protein; 289 AA.

AC AAB44522;

DT 08-FEB-2001 (first entry)

DE Virulence gene protein #2.

XX Virulence gene; antibacterial; vaccine; bacterial infection; septicemia;

XX Virulence gene; antibacterial; vaccine; bacterial infection; septicemia;

XX Pasteurella multocida.

XX WC200061724-A2.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US009218.

XX 09-APR-1999; 99US-0128689P.

XX 10-SEP-1999; 99US-0153453P.

XX (PHAA ) PHARMACIA & UPOJOHN INC.

XX Lowery DE, Fuller TE, Kennedy MJ;

XX WPI; 2000-647422/62.

XX N-PSDB; AAC79582.

XX Attenuated Pasteurellaceae bacteria comprising mutations in virulence genes, useful as a live attenuated vaccine against bacterial infections.

XX Claim 39; Page 68-69; 322pp; English.

XX The family Pasteurellaceae encompasses several pathogens that infect a wide variety of animals. The present invention relates to virulence genes from Pasteurellaceae. The present sequence is a protein encoded by one such virulence gene. The virulence genes of the present invention may be mutated in order to produce an inactive gene. The inactive virulence gene may in turn be used to produce a vaccine, which is useful for treating bacterial infections such as septicemias, bronchopneumonias, rhinitis and wound infections

XX Sequence 289 AA;

Query Match 100.0%; Score 1429; DB 3; Length 289;



```

Best Local Similarity 100.0%; Pred. No. 5.7e-121;
Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAGAKAIRKIASVSKTKIKAMENWAASKORKTQERMSSRPSETIRNVIHSVKAT 60
1 MAGAKAIRKIASVSKTKIKAMENWAASKORKTQERMSSRPSETIRNVIHSVKAT 60
61 IGYKHPFLVDREVKKVGMIVSTDRLGCGGLNVNLFKTVLNEMKEWKEDVSQLSLIGS 120
61 IGYKHPFLVDREVKKVGMIVSTDRLGCGGLNVNLFKTVLNEMKEWKEDVSQLSLIGS 120
121 KSNFPQSLGILIKIQDSGIGDTPSVQLIGSVNSMIDAYKKGEVDVVLVYKFNPTMS 180
121 KSNFPQSLGILIKIQDSGIGDTPSVQLIGSVNSMIDAYKKGEVDVVLVYKFNPTMS 180
181 QKPVLEKLIPLPELNDNDELBERKQWDIYVEPPAKVLLDNLLVRYLESQVQAAVENLAS 240
181 QKPVLEKLIPLPELNDNDELBERKQWDIYVEPPAKVLLDNLLVRYLESQVQAAVENLAS 240
241 EQAARVAMKAATDNAGNLINELQLVNFKARQASITNELNEIVAGAAAI 289
241 EQAARVAMKAATDNAGNLINELQLVNFKARQASITNELNEIVAGAAAI 289

SULT 2
P54474
ABP54474 standard; protein; 289 AA.

```

SULT 2  
P54474  
ABP54474 standard; protein; 289 AA.

ABP54474;  
06-AUG-2003 (revised)  
24-JAN-2003 (first entry)  
Pasteurella multocoda atpG protein SEQ ID NO:4.  
Antibacterial; vaccine; gram negative bacterial virulence gene;  
identification; virulence; Pasteurellaceae.  
Pasteurella multocida.  
WO200275507-A2.  
26-SEP-2002.  
17-JAN-2002; 2002WO-US001971.  
15-MAR-2001; 2001US-00809665.  
(PHAA ) PHARMACIA & UPJOHN CO.  
Lowery DE, Fuller TE, Kennedy MJ;  
WPI; 2002-740868/80.  
N-PSDB; ABQ83459.  
New mutant gram-negative bacteria, useful as vaccines and for identifying  
new anti-bacterial agents that target virulence genes and their products.  
Claim 36; Page 73-74; 350pp; English.  
The present invention describes a gram-negative bacteria comprising a  
mutation in a gene, where the mutation results in decreased activity of a  
gene product encoded by the mutated gene. Also described is a method for  
producing a gram-negative bacteria mutant or an attenuated  
Pasteurellaceae bacteria. The mutated genes have antibacterial activity  
and can be used in vaccines. The gram-negative bacteria or the attenuated  
Pasteurellaceae bacteria can be used as vaccines in the fields of human  
medicine or veterinary medicine, and for identifying new antibacterial  
agents that target the virulence genes and their products. ABQ83458 to  
ABQ83578 and ABP54473 to ABP54551 represents sequences used in the  
exemplification of the present invention.. (Updated on 06-AUG-2003 to  
correct OS field.)

```

SQ Sequence 289 AA;
Query Match      100.0%; Score 1429; DB 5; Length 289;
Best Local Similarity 100.0%; Pred. No. 5.7e-121;
Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  MAGAKEIRTKIASVKSTQKIITKAMEWVAASKRKTKTQERMSSSRPYSETIRNVISHVSKAT 60
Db   1  MAGAKEIRTKIASVKSTQKIITKAMEWVAASKRKTKTQERMSSSRPYSETIRNVISHVSKAT 60

QY  61  IGYKHPPFLVDREVKKVGMIVVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQLSLIGS 120
Db   61  IGYKHPPFLVDREVKKVGMIVVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQLSLIGS 120

QY  121  KSNFFQSLGKILITQDSGIGDTPSVQLIGSVNSMIDAYKKGEVDVYVLVTNKFINTMS 180
Db   121  KSNFFQSLGKILITQDSGIGDTPSVQLIGSVNSMIDAYKKGEVDVYVLVTNKFINTMS 180

QY  181  QKPVLKLIPIPELNDDELGERKQWDYIYEPDAKVLNDLLVRYLESOVYQAAVENVLAS 240
Db   181  QKPVLKLIPIPELNDDELGERKQWDYIYEPDAKVLNDLLVRYLESOVYQAAVENVLAS 240

QY  241  EQAARWVAMKAATDNAGNLINELQLVTNKAQASITNELNEIVAGAAA 289
Db   241  EQAARWVAMKAATDNAGNLINELQLVTNKAQASITNELNEIVAGAAA 289

RESULT 3
AAB44579
ID  AAB44579 standard; protein; 288 AA

```

RESULT 3  
 AAB44579  
 ID AAB44579 standard; protein; 288 AA.  
 XX  
 AC AAB44579;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Virulence gene protein #59.  
 XX  
 KW Virulence gene; antibacterial; vaccine; bacterial infection; septicemia;  
 XX bronchopneumonia; rhinitis; wound infection.  
 XX  
 OS Actinobacillus pleuropneumoniae.  
 XX  
 PN WC200061724-A2.  
 XX  
 PD 19-OCT-2000.  
 XX  
 PF 06-APR-2000; 2000WO-US009218.  
 XX  
 PR 09-APR-1999; 99US-0128689P.  
 XX 10-SEP-1999; 99US-0153453P.  
 XX (PHAA ) PHARMACIA & UPJOHN INC.  
 XX  
 Lowery DE, Fuller TE, Kennedy MJ;  
 WPI; 2000-647422/62.  
 N-PSDB; AAC79654.  
 XX  
 Attenuated Pasteurellaceae bacteria comprising mutations in virulence  
 genes, useful as a live attenuated vaccine against bacterial infections.  
 Claim 39; Page 292; 322pp; English.  
 XX  
 The family Pasteurellaceae encompasses several pathogens that infect a  
 wide variety of animals. The present invention relates to virulence genes  
 from Pasteurellaceae. The present sequence is a protein encoded by one  
 such virulence gene. The virulence genes of the present invention may be  
 mutated in order to produce an inactive gene. The inactive virulence gene  
 may in turn be used to produce a vaccine, which is useful for treating  
 bacterial infections such as septicemias, bronchopneumonias, rhinitis and  
 wound infections  
 XX  
 Sequence 288 AA:

Query Match 79.3%; Score 1132.5; DB 3; Length 288;  
 Best Local Similarity 74.4%; Pred. No. 4.4e-94;  
 Matches 215; Conservative 45; Mismatches 28; Indels 1; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASMKRTOERMSSRPYSETIRNVISHVSKAT 60  
 DB 1 MAGAKEIRTKIASVKSTOKITKAMEMVATSMKRTQERMAASRPYSETIRKVIISHIAKGS 60

QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLFKTVLNEKWKEDVSVQSLIGS 120  
 DB 61 IGYKHPFLTERDIKKVGLVIVVSTDRGLCGGLNINLFKATLNEFKTWKDKDVSVELGLVGS 120

QY 121 KSINFFQSLGKILTDOSGIDTPSVQELIGSVNSMIDAYKKEGVVVYLVNKFINTMS 180  
 DB 121 KGVSYFQNLGLNVRQVTLGDNPEMERIVGAVNEMINAFRNGEVDVAVYVNRFEPTMS 180

QY 181 QKPVLEKLIPLDELNDLGERKQVWDYIYEPDAKVLNLLVRYLESQVQAAVENLAS 240  
 DB 181 QKPVIAQLLEPLKLDDEL-DTKGSDYIYEPNPQVLLDLSLLVRYLETQVQAVVDNLAS 239

QY 241 EQAARVMVAKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAAI 289  
 DB 240 EQAARVMVAKAATDNAGTLIDELQLVYNKARQASITNELNEIVAGAAAI 288

RESULT 4  
 ABP54531  
 ID ABP54531 standard; protein; 288 AA.  
 AC ABP54531;  
 DT 24-JAN-2003 (first entry)  
 DE Actinobacillus pleuropneumoniae atpG protein SEQ ID NO:133.  
 KW Antibacterial; vaccine; gram negative bacterial virulence gene;  
 KW identification; virulence; Pasteurellaceae.  
 OS Actinobacillus pleuropneumoniae.  
 XX WO200275507-A2.  
 PD 26-SEP-2002.  
 PF 17-JAN-2002; 2002WO-US001971.  
 PR 15-MAR-2001; 2001US-00809665.  
 PA (PHAA ) PHARMACIA & UPJOHN CO.  
 PI Lowery DE, Fuller TE, Kennedy MJ;  
 DR WPI; 2002-740868/80.  
 DR N-PSDB; ABQ83531.  
 PT New mutant gram-negative bacteria, useful as vaccines and for identifying  
 PT new anti-bacterial agents that target virulence genes and their products.  
 PS Claim 36; Page 298-299; 350pp; English.

The present invention describes a gram-negative bacteria comprising a mutation in a gene, where the mutation results in decreased activity of a gene product encoded by the mutated gene. Also described is a method for producing a gram-negative bacteria mutant or an attenuated Pasteurellaceae bacteria. The mutated genes have antibacterial activity and can be used in vaccines. The gram-negative bacteria or the attenuated Pasteurellaceae bacteria can be used as vaccines in the fields of human medicine or veterinary medicine, and for identifying new antibacterial agents that target the virulence genes and their products. ABQ83458 to ABQ83578 and ABP54473 to ABP54551 represents sequences used in the exemplification of the present invention

Sequence 288 AA;  
 Query Match 79.3%; Score 1132.5; DB 5; Length 288;  
 Best Local Similarity 74.4%; Pred. No. 4.4e-94;  
 Matches 215; Conservative 45; Mismatches 28; Indels 1; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASMKRTOERMSSRPYSETIRNVISHVSKAT 60  
 DB 1 MAGAKEIRTKIASVKSTOKITKAMEMVATSMKRTQERMAASRPYSETIRKVIISHIAKGS 60

QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLFKTVLNEKWKEDVSVQSLIGS 120  
 DB 61 IGYKHPFLTERDIKKVGLVIVVSTDRGLCGGLNINLFKATLNEFKTWKDKDVSVELGLVGS 120

QY 121 KSINFFQSLGKILTDOSGIDTPSVQELIGSVNSMIDAYKKEGVVVYLVNKFINTMS 180  
 DB 121 KGVSYFQNLGLNVRQVTLGDNPEMERIVGAVNEMINAFRNGEVDVAVYVNRFEPTMS 180

QY 181 QKPVLEKLIPLDELNDLGERKQVWDYIYEPDAKVLNLLVRYLESQVQAAVENLAS 240  
 DB 181 QKPVIAQLLEPLKLDDEL-DTKGSDYIYEPNPQVLLDLSLLVRYLETQVQAVVDNLAS 239

QY 241 EQAARVMVAKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAAI 289  
 DB 240 EQAARVMVAKAATDNAGTLIDELQLVYNKARQASITNELNEIVAGAAAI 288

RESULT 5  
 ABP54547  
 ID ABP54547 standard; protein; 288 AA.  
 AC ABP54547;  
 DT 29-AUG-2003 (revised)  
 DT 24-JAN-2003 (first entry)  
 DE Pasteurella haemolytica atpG protein SEQ ID NO:167.  
 KW Antibacterial; vaccine; gram negative bacterial virulence gene;  
 KW identification; virulence; Pasteurellaceae.  
 OS Mannheimia haemolytica.  
 XX WO200275507-A2.  
 PD 26-SEP-2002.  
 PF 17-JAN-2002; 2002WO-US001971.  
 PR 15-MAR-2001; 2001US-00809665.  
 PA (PHAA ) PHARMACIA & UPJOHN CO.  
 PI Lowery DE, Fuller TE, Kennedy MJ;  
 DR WPI; 2002-740868/80.  
 DR N-PSDB; ABQ83549.  
 PT New mutant gram-negative bacteria, useful as vaccines and for identifying  
 PT new anti-bacterial agents that target virulence genes and their products.  
 PS Claim 36; Page 330-331; 350pp; English.

The present invention describes a gram-negative bacteria comprising a mutation in a gene, where the mutation results in decreased activity of a gene product encoded by the mutated gene. Also described is a method for producing a gram-negative bacteria mutant or an attenuated Pasteurellaceae bacteria. The mutated genes have antibacterial activity and can be used in vaccines. The gram-negative bacteria or the attenuated Pasteurellaceae bacteria can be used as vaccines in the fields of human medicine or veterinary medicine, and for identifying new antibacterial agents that target the virulence genes and their products. ABQ83458 to ABQ83578 and ABP54473 to ABP54551 represents sequences used in the



XX New *Proteus mirabilis* polypeptides and polynucleotides, useful as  
 PT reagents for diagnosis of bacterial disease, as components of  
 PT antibacterial vaccines, as targets for antibacterial drugs, or as  
 PT biocontrol agents for plants.  
 XX  
 FS Disclosure; SEQ ID NO 7449; 870pp; English.  
 XX  
 XX The invention relates to new *Proteus mirabilis* polypeptides and  
 CC polynucleotides. The invention also relates to antibodies against the  
 CC polypeptides, methods for producing the polypeptides, a method of  
 CC generating vaccines for immunising an individual against *P. mirabilis*, a  
 CC method for evaluating a compound for the ability to bind a *P. mirabilis*  
 CC polypeptide and a method for screening test compounds for anti-bacterial  
 CC activity. The polypeptides and polynucleotides are useful as molecular  
 CC targets for diagnosing, preventing and treating pathological conditions  
 CC resulting from bacterial infection, as reagents for diagnosis of  
 CC bacterial diseases, as components of antibacterial vaccines, as targets  
 CC for antibacterial drugs or as bio-control agents for plants. This  
 CC sequence represents a *Proteus mirabilis* polypeptide of the invention.  
 XX  
 SQ Sequence 291 AA;

Query Match 70.3%; Score 1004; DB 7; Length 291;  
 Best Local Similarity 67.1%; Pred. No. 2e-82;  
 Matches 194; Conservative 49; Mismatches 44; Indels 2; Gaps 1;  
 QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASRKQTKQERMSSSRPYSETIRNVISHVSKAT 60  
 DB 5 MAGAKEIRSKIASVQNTQKITKAMEMVAASRKQTKQERMAASRPYATETMSVIGHALGN 64  
 QY 61 IGYKHPFLVDREVKVGMIVSTDRGLCGGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 120  
 DB 65 LEYKHPYLEDRVKRGVYLVSTDRGLCGGLNINLFKLLADMKWSDKGVEVDLALGS 124  
 QY 121 KSNFFQSLGKILTDQSGIGDTPSVQELIGSVNMDIAYKKGVEDVYVLYNKFINTMS 180  
 DB 125 KAVSFFASVGGNVGVQVTGMDDPQLSDLLGPNVIMLQAYDEGRDLKLYVANKFINTMA 184  
 QY 181 QKPVLEKLIPLPELDNDELGERKQVWDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
 DB 185 QEPKILQVLPPLPGDDELKEKS--WDYLYEPDPKTLTLLRRYTESQVYQSVVENLAS 242  
 QY 241 EQAARWAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289  
 DB 243 EQAARWAMKAATDNGSLIKELQLVYNKARQASITQELTEIVSGAAAV 291

RESULT 8  
 AAG98289  
 ID AAG98289 standard; protein; 287 AA.  
 XX  
 AC AAG98289;  
 XX  
 DT 21-SEP-2001 (first entry)  
 XX  
 DE Escherichia coli protein sequence SEQ ID NO:337.  
 XX  
 CW Escherichia coli; identification; proliferation; microorganism;  
 CW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;  
 CW bacterial growth inhibition.  
 XX  
 DE Escherichia coli.  
 XX  
 PN W0200148209-A2.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 19-DEC-2000; 2000WO-US034419.  
 XX  
 PR 23-DEC-1999; 99US-0173005P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.

XX  
 PT  
 XX  
 DR Forsyth RA, Ohlsen KL, Zyskind JW;  
 DR WPI: 2001-457376/49.  
 DR N-PSDB; AAH81345.  
 XX  
 PT Novel nucleic acids encoding proteins required for *Escherichia coli*  
 PT proliferation, useful for screening for antimicrobial agents.  
 XX  
 XX Claim 19; Page 468; 596pp; English.  
 XX  
 CC The present invention describes a purified or isolated nucleic acid  
 CC sequence (I) consisting essentially of one of the 93 nucleotide sequences  
 CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a  
 CC microorganism is capable of inhibiting proliferation of a microorganism.  
 CC (II) have antibacterial and antibiotic activities, and can be used in gene  
 CC therapy. Expression of (I) in a microorganism inhibits proliferation of  
 CC the microorganism, and the manufactured antibiotic is useful for reducing  
 CC the activity or level of a gene product required for proliferation of a  
 CC microorganism in a subject, specifically humans. The nucleic acids that  
 CC inhibit bacterial growth or proliferation can be used as antisense  
 CC therapeutics for killing bacteria. In addition to therapeutic  
 CC applications, the nucleic acid sequences complementary to sequences  
 CC required for proliferation can be used as diagnostic tools. For example,  
 CC nucleic acid probes complementary to proliferation-required sequences  
 CC that are specific for particular species of microorganisms can be used as  
 CC probes to identify particular microorganism species in clinical  
 CC specimens. AAH81295 to AAH81487 encode the *Escherichia coli* proteins  
 CC given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent  
 CC oligonucleotides, which are used in the exemplification of the present  
 CC invention  
 XX  
 SQ Sequence 287 AA;

Query Match 69.3%; Score 991; DB 4; Length 287;  
 Best Local Similarity 56.1%; Pred. No. 3e-81;  
 Matches 191; Conservative 47; Mismatches 49; Indels 2; Gaps 1;  
 QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASRKQTKQERMSSSRPYSETIRNVISHVSKAT 60  
 DB 1 MAGAKEIRSKIASVQNTQKITKAMEMVAASRKQTKQERMAASRPYATETMSVIGHALGN 60  
 QY 61 IGYKHPFLVDREVKVGMIVSTDRGLCGGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 120  
 DB 61 LEYKHPYLEDRVKRGVYLVSTDRGLCGGLNINLFKLLAEKMTWTDKVQCDLALIGS 120  
 QY 121 KSNFFQSLGKILTDQSGIGDTPSVQELIGSVNMDIAYKKGVEDVYVLYNKFINTMS 180  
 DB 121 KGVSFNSVGGNVGVQVTGMGNPSELIGFVKYMLQAYDEGRDLKLYVSNKFINTMS 180  
 QY 181 QKPVLEKLIPLPELDNDELGERKQVWDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
 DB 181 QVPTISQLPLPASDDDL--KHKSWDYLYEPDPKALLDTLLRRYTESQVYQGVVENLAS 238  
 QY 241 EQAARWAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289  
 DB 239 EQAARWAMKAATDNGSLIKELQLVYNKARQASITQELTEIVSGAAAV 287

RESULT 9  
 ABO64099  
 ID ABO64099 standard; protein; 297 AA.  
 XX  
 AC ABO64099;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Klebsiella pneumoniae polypeptide seqid 10616.  
 XX  
 KW Recombinant expression vector; transcription regulatory element;  
 KW Klebsiella pneumoniae protein; antibacterial; Vaccine.  
 XX  
 OS Klebsiella pneumoniae.

```

XX PN US6610836-B1.
XX PD
XX PD 26-AUG-2003.
XX PF 27-JAN-2000; 2000US-00489039.
XX PR 29-JAN-1999; 99US-0117747P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL, Osborne M;
XX PI N-PSDB; ACH97650.
XX DR WPI: 2003-895346/82.
XX DR N-PSDB; ACH97650.
XX PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX PT preparing a vaccine composition against Klebsiella pneumoniae.
XX PS Disclosure; SEQ ID NO 10616; 932pp; English.
XX CC The invention describes a new isolated nucleic acid encoding a Klebsiella
XX CC pneumoniae polypeptide. Also described are: a recombinant expression
XX CC vector comprising the nucleic acid, operably linked to a transcription
XX CC regulatory element; and a cell comprising the recombinant expression
XX CC vector. The nucleic acid is useful for preparing a vaccine composition
XX CC against Klebsiella pneumoniae. This is the amino acid sequence of a
XX CC Klebsiella pneumoniae polypeptide of the invention
XX SQ Sequence 297 AA;
Query Match 68.4%; Score 978; DB 7; Length 297;
Best Local Similarity 65.1%; Pred. No. 4.7e-80;
Matches 188; Conservative 50; Mismatches 49; Indels 2; Gaps 1;
QY 1 MAGAKEIRTKIASVSTQKITKAMEMVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60
DB 11 MAGAKEIRTKIASVSTQKITKAMEMVAASQKRSQERMAASRPYADTKRKVIGHLAN 70
QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNNLFTVLNEMKEWKEKDVSVQLSIG 120
DB 71 LEYKHPFLVDREVKKVGMIVVSTDRGLCGGLNNLFTVLNEMKEWKEKDVSVQLSIG 130
QY 121 KSNFFQSLGKILQDSGIGDTPSVQLIGSVNSMIDAYKKGEVDVVLVYVKNKFIN 180
DB 131 KGVSPFNSVGVNVAQVTGMDNPSLSELPVKKVLMQAYDEGRDLKLYVSNKFIN 190
QY 181 QKPVLEKLIPLPELDNDELGERKQVWDYIYEPDAKVLNLLVRYLESQVYQAAVEN 240
DB 191 QVPTITQLPLPASEDADL--KRSWDYLYEPDPKALLDTLLRRYVESQVYQGVVEN 248
QY 241 EQAARMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAI 289
DB 249 EQAARMVAMKAATDNGSLIKEQLVYNKARQASITQELTEIVSGAAV 297
RESULT 10
ABO70334
ID ABO70334 standard; protein; 291 AA.
XX AC ABO70334;
XX AC
XX DT 29-JUL-2004 (first entry)
XX DE Pseudomonas aeruginosa polypeptide #2509.
XX DE Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
XX OS Pseudomonas aeruginosa.
XX PN US6551795-B1.
XX PD 22-APR-2003.

```

```

XX 18-FEB-1999; 99US-00252991.
XX 18-FEB-1999; 98US-0074788P.
XX 27-JUL-1998; 98US-0094130P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI: 2003-615309/58.
XX DR N-PSDB; ABD03905.
XX PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of
XX PT pathological conditions resulting from bacterial infection.
XX PS Disclosure; SEQ ID NO 19080; 455pp; English.
XX CC The invention relates to Pseudomonas aeruginosa polypeptides and the
XX CC polynucleotides encoding them. The sequences are useful in diagnosis and
XX CC therapy of pathological conditions, as molecular targets for diagnostics,
XX CC prophylaxis and treatment of pathological conditions resulting from a
XX CC bacterial infection, for evaluating a compound, such as a polypeptide,
XX CC for the ability to bind a P. aeruginosa nucleic acid, as components of
XX CC effective antibacterial targets, as targets for antibacterial drugs,
XX CC including anti-P. aeruginosa drugs, as templates for recombinant
XX CC production of P. aeruginosa-derived peptides or polypeptides, as target
XX CC components for diagnosis and/or treatment of P. aeruginosa-caused
XX CC infection, and in detection of P. aeruginosa sequences or other sequences
XX CC of Pseudomonas species using biochip technology. Sequences ABO67826-
XX CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification but was obtained in electronic format from USPTO at
XX CC seqdata.uspto.gov/sequence.html
XX SQ Sequence 291 AA;
Query Match 64.8%; Score 926.5; DB 7; Length 291;
Best Local Similarity 62.3%; Pred. No. 2.1e-75;
Matches 180; Conservative 50; Mismatches 56; Indels 3; Gaps 1;
QY 1 MAGAKEIRTKIASVSTQKITKAMEMVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60
DB 6 MAGAKEIRTKIASVSTQKITNAMEKVAVSKVKAQMEVACRPYAEIRIQVIGHLAN 65
QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNNLFTVLNEMKEWKEKDVSVQLSIG 120
DB 66 PEYRHFPMVEREVRVGIYVSSDRGLCGGLNNLFTVLNEMKEWKEKDVSVQLSIG 125
QY 121 KSNFFQSLGKILQDSGIGDTPSVQLIGSVNSMIDAYKKGEVDVVLVYVKNKFIN 180
DB 126 KGASFRSFGNVAASHLGEPSINDLIGSVKWLDAYLEGRIDRLFVVSNFVNT 185
QY 181 QKPVLEKLIPLPELDNDELGERKQVWDYIYEPDAKVLNLLVRYLESQVYQAAVEN 240
DB 186 QKPTVEQLIPLVADDDQEL--KHHWDYLYEPDAKSLDGLLVRYVESQVYQAVVEN 242
QY 241 EQAARMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAI 289
DB 243 EQAARMVAMKAATDNAGELISDQLIYNKARQAITQELTEIVSGAAV 291
RESULT 11
ADA33607
ID ADA33607 standard; protein; 292 AA.
XX AC ADA33607;
XX AC
XX DT 20-NOV-2003 (first entry)
XX DE Acinetobacter baumannii protein #768.
XX

```



XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX PI WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS88668.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 54840; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological actions. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 969 AA;  
 Query Match 54.2%; Score 774; DB 4; Length 969;  
 Best Local Similarity 55.2%; Pred. No. 7.9e-61;  
 Matches 155; Conservative 48; Mismatches 70; Indels 8; Gaps 3;  
 QY 7 IRTKASVSTQKITKAMEMVAASKVRK--TQRMSSSRPYSETIRNVISHVSKATGYK 64  
 Db 672 IRTSCTERTGTQTRWPGRLLESDSFPHGLSQE----SYPAETMRKVIGHLAGHLEYK 727  
 QY 65 HPLVDREVKYGMIVSTDRGLCGLVNLFKTVLNEMKEWKEKDVSVQLSLGSKIN 124  
 Db 728 HPVLEDRDVKRVGYLVSTDRGLCGLVNLFKLLAEMKWTBQGVQCDLAMIGSKVS 787  
 QY 125 FQSLGKILTDQSGIGDTPPSVEQLIGSVNSMIDAYKKEGVDDVYLVYNKFTINTMSQKV 184  
 Db 788 FNSVGNVVAQVTGMDGNPSLSLIGSVKVMQAYDEVRLDKLYICNHQFITPMSQVPT 847  
 QY 185 LEKLPLPELDNDELGERQVWDYIYEPDAKVLNLLVRYLESQVYQAQAVENLASEQAA 244  
 Db 848 ISQLPLPASDDDDJ--KHKSWDYIYEPDPKALLDTLLRRVSEQVTVGVVENLASEQAA 905  
 QY 245 RVAMKAATDNAGNLINELQLVYNKARQASITNELNEIVAG 285  
 Db 906 RVAMKAATDNGSLIKELQLVYNKARQASITQELTEIVSG 946

RESULT 14  
 ABP77380  
 ID ABP77380 standard; protein; 291 AA.  
 XX  
 AC ABP77380;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae amino acid sequence SEQ ID 1290.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 FN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB002069.  
 XX  
 PR 12-FEB-2001; 2001GB-00003424.  
 XX  
 PA (CHIR-) CHIRON SPÅ.  
 XX  
 PI Pontana MR, Pizza M, Masignani V, Monaci E;  
 XX  
 DR WPI; 2003-058415/05.  
 DR N-PSDB; AB238350.  
 XX  
 CC New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 CC medicament for treating or preventing N. gonorrhoeae infection.  
 XX  
 PS Disclosure; Page 281; 815pp; English.  
 XX  
 CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in the invention ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention  
 XX  
 SQ Sequence 291 AA;  
 Query Match 53.2%; Score 760; DB 6; Length 291;  
 Best Local Similarity 51.5%; Pred. No. 2.7e-60;  
 Matches 150; Conservative 64; Mismatches 75; Indels 2; Gaps 2;  
 QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASKVRKKTQERMSSSRPYSETIRNVISHVSKAT 60  
 Db 1 MAVGKEILTIRSVQNTQKITKAMQVSTSKVRKTQERMSSSRPYSETIRNVISHVSKAT 60  
 QY 61 IGYKHPFLVD-REVKKYGMIVSTDRGLCGLVNLFKTVLNEMKEWKEKDVSVQLSLG 119  
 Db 61 TDHGTPLESREIRRVGFILITSDKGLCGLVNLFKTVLNEMKEWKEKDVSVQLSLG 120  
 QY 120 SKSINFFQSLGKILTDQSGIGDTPPSVEQLIGSVNSMIDAYKKEGVDDVYLVYNKFTINTM 179  
 Db 121 SKGLMACQSIGLVNVASAVNLGDTPKVEMLLGLPTLTFQRYEKHEIDRIHLVYSGFVNTM 180  
 QY 180 SQKPVLEKILPLPE-LDNDELGERQVWDYIYEPDAKVLNLLVRYLESQVYQAQAVENL 238  
 Db 181 RQERMEVLIPIGENVIGDSAPKPSFWEYRYEPTALVLELVRRYLESVYQALSNNM 240  
 QY 239 ASEQAARVAMKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAAI 289  
 Db 241 ASEQAARVAMKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAAV 291  
 RESULT 15  
 ABG13826  
 ID ABG13826 standard; protein; 924 AA.

XX	AC	ABG13826;	
XX	DT	18-FEB-2002 (first entry)	
XX	DE	Novel human diagnostic protein #13817.	
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX	OS	Homo sapiens.	
XX	FN	W0200175067-A2.	
XX	PD	11-OCT-2001.	
XX	FF	30-MAR-2001; 2001WO-US009631.	
XX	PR	31-MAR-2000; 2000US-00540217.	
XX	PR	23-AUG-2000; 2000US-00649167.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	FI	Drmanac RT, Liu C, Tang YT;	
XX	DR	WPI; 2001-639362/73.	
XX	DR	N-PSDB; AAS78013.	
XX	PT	New isolated polynucleotide and encoded polypeptides, useful in	
XX	PT	diagnostics, forensic, gene mapping, identification of mutations	
XX	PT	responsible for genetic disorders or other traits and to assess	
XX	XX	biodiversity.	
XX	PS	Claim 20; SEQ ID NO 44185; 103pp; English.	
XX	CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)	
XX	CC	sequences. (I) is useful as hybridisation probes, polymerase chain	
XX	CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,	
XX	CC	and in recombinant production of (II). The polynucleotides are also used	
XX	CC	in diagnostics as expressed sequence tags for identifying expressed	
XX	CC	genes. (II) is useful in gene therapy techniques to restore normal	
XX	CC	activity of (II) or to treat disease states involving (II). (II) is	
XX	CC	useful for generating antibodies against it, detecting or quantitating a	
XX	CC	polypeptide in tissue, as molecular weight markers and as a food	
XX	CC	supplement. (II) and its binding partners are useful in medical imaging	
XX	CC	of sites expressing (II). (I) and (II) are useful for treating disorders	
XX	CC	involving aberrant protein expression or biological activity. The	
XX	CC	polypeptide and polynucleotide sequences have applications in	
XX	CC	diagnostics, forensics, gene mapping, identification of mutations	
XX	CC	responsible for genetic disorders or other traits to assess biodiversity	
XX	CC	and to produce other types of data and products dependent on DNA and	
XX	CC	amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic	
XX	CC	amino acid sequences of the invention. Note: The sequence data for this	
XX	CC	patent did not appear in the printed specification, but was obtained in	
XX	CC	electronic format directly from WIPO at	
XX	CC	ftp.wipo.int/pub/published_pct_sequences	
XX	XX		
XX	SQ	Sequence 924 AA;	
		Query Match 45.0%; Score 642.5; DB 4; Length 924;	
		Best Local Similarity 52.9%; Pred. No. 6.2e-49;	
		Matches 129; Conservative 30; Mismatches 38; Indels 47; Gaps 2	
QY		42 SRPYETTRNVISHVSKATICKPEFLVDREVKVGVMIVSTDRGICGGINLVNLFKTVLN 101	
Db		705 SYPYAEYTKRWKIGHUAGHNLSYKPIYLEDKRVKGYLVVSTDRGICGGINLVNLFKGLIA 764	
QY		102 EMKEWKEKDVSVQLSGLSKSINFFQSGLIGIKILTQDSGIGDTFSVEQLIGSVNSMIDAYK 161	
Db		765 EMKTTWDKGVQCDLAMISSKGVSFNSVG----- 793	
QY		162 KGEVDVTVLVNKNKINTMSQPVLEKLIPILPELDNDELGERKQVDYIYEPDAKVLNLL 221	



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2004, 09:46:43 ; Search time 16 seconds  
(without alignments)  
1737.915 Million cell updates/sec

Title: US-09-545-199f-4  
Perfect score: 1429  
Sequence: 1 MAGAKEIRTKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_79:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1173	82.1	289	1 E64071	H+-transporting tw
2	998.5	69.9	288	2 G82036	ATP synthase Fl, g
3	991	69.3	287	1 PWECG	H+-transporting tw
4	991	69.3	287	2 C91213	membrane-bound ATP
5	991	69.3	287	2 D86059	membrane-bound ATP
6	989	69.2	287	2 AF0500	H+-transporting tw
7	978	68.4	287	2 AB0954	ATP synthase gamma
8	976.5	68.3	288	1 S06081	H+-transporting tw
9	926.5	64.8	286	2 D82952	ATP synthase gamma
10	808	56.5	290	2 G84930	H+-transporting tw
11	795	55.6	287	2 H82715	ATP synthase gamma
12	754	52.8	291	2 B81970	ATP synthase, gamm
13	753	52.7	291	2 H81024	H+-transporting tw
14	566.5	39.6	285	2 G31482	ATP synthase Fl, g
15	536.5	37.5	285	2 C84119	H+-transporting tw
16	533.5	37.3	287	2 C69592	ATP synthase gamma
17	523.5	36.6	289	2 B86845	H+-transporting tw
18	519.5	36.4	289	2 S17725	H+-transporting tw
19	509.5	35.7	315	2 AD1807	ATP synthase chain
20	508.5	35.6	315	2 H31050	H+-transporting tw
21	505.5	35.4	286	2 S01402	H+-transporting tw
22	504.5	35.3	314	1 FWBYG	H+-transporting tw
23	503.5	35.2	316	1 PWYCG	H+-transporting tw
24	501.5	35.1	292	2 G95175	ATP synthase Fl, g
25	501.5	35.1	292	2 H98041	H+-transporting tw
26	496.5	34.7	282	2 S28040	Na+-transporting A
27	495.5	34.7	315	2 S36979	H+-transporting tw
28	494.5	34.6	315	2 S32401	H+-transporting tw
29	492	34.4	290	2 AB1391	H+-transporting AT

30	492	34.4	290	2 AD1766	H+-transporting AT
31	490	34.3	288	2 E90003	ATP synthase gamma
32	487	34.1	292	2 JC5740	membrane-bound pro
33	477.5	33.4	311	2 S49845	H+-transporting tw
34	476.5	33.3	305	2 A07075	probable atpG prot
35	471.5	33.0	301	2 E64661	H+-transporting tw
36	468	32.8	299	1 PQQFG	H+-transporting tw
37	466.5	32.6	301	2 E71855	ATP synthase Fl, c
38	460	32.2	294	2 B81427	H+-transporting tw
39	459.5	32.2	292	2 G97673	H+-transporting tw
40	459.5	32.2	292	2 AF2898	ATP synthase gamma
41	458	32.1	377	1 PNTNG	H+-transporting tw
42	455	31.8	364	1 PWSFG	H+-transporting tw
43	454	31.8	373	2 B39732	H+-transporting tw
44	449.5	31.5	286	2 S04673	H+-transporting tw
45	449.5	31.5	386	2 A39732	H+-transporting tw

## ALIGNMENTS

### RESULT 1

E64071  
H+-transporting two-sector ATPase (EC 3.6.3.14) gamma chain - Haemophilus influenzae (st  
C/Species: Haemophilus influenzae  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: E64071  
R.Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Maidman, J  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, ,  
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A/Reference number: A64000; MUID:95350630; PMID:7542800  
A/Accession: E64071  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-289 <TIGR>  
A/Cross-references: UNIPROT:P43716; GB:U32730; GB:L42023; NID:G3212191; PIDN:AAC22138.1;  
C/Superfamily: H+-transporting ATP synthase gamma chain  
C/Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match	82.1%	Score 1173;	DB 1;	Length 289;
Best Local Similarity	76.5%;	Pred. No. 1.5e-69;		
Matches	221;	Conservative 42;	Mismatches 26;	Indels 0; Gaps 0;
QY	1	MAGAKEIRTKIASVKSTQKITKAMWVAASXMRKTOERMSRRPYSETIRNVISHVSKAT	60	
Db	1	MAGAKEIRTKIASVQSTQKITKAMWVAATSKVRKTDRAASRPYSETIRNVISHVSKAS	60	
QY	61	IGYKHPFLVDREVKKVGMIVSTDRGLCGGLNVNLFKTVLNMKEWKEKDVSVQLSLIGS	120	
Db	61	IGYKHPFLVEREYKKGILVISTDRGCGGLNVNLFKTTLNQIKWKKEQNTSTDGLIGS	120	
QY	121	KSINFFQSIGIKILTODSGIGDTPTSEVQLIGSVNMDIAYKKGVEDVYVLYNKFINTMS	180	
Db	121	KGISFRSFGFNKQGLSGLGTPALEELIGVANTMFAYRNGEIDAVYIAYNKFVNTMS	180	
QY	181	QKPVLEKLPLPELDNDELGERKQVWDYIYEPDAKVLNLDNLVRYLESQVYQAAVENLAS	240	
Db	181	QKPVQQLVPLPESKDDHLNEQQTWDYLYPEPKVKLLDSLLVRYLESQIYQAVDVNAS	240	
QY	241	EQAARMVAKAATDAGNLINELQVYNKARQASITNELNEIVAGAAAI	289	
Db	241	EQAARMVAKAATDAGNLINDLRLVYNKARQASITNELNEIVAGAAAI	289	

### RESULT 2

G82036  
ATP synthase Fl, gamma chain VC2765 [imported] - Vibrio cholerae (strain N16961 serogrou  
C/Species: Vibrio cholerae  
C/Date: 19-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C/Accession: G82036

R; Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, J.; R.A.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A; Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A; Reference number: A82035; MUID:20406833; PMID:10952301  
 A; Accession: G82036  
 A; Status: preliminary  
 A; Molecule type: DNA  
 A; Residues: 1-288 <HEI>  
 A; Cross-references: UNIPROT:Q9NMH4; GB:AE004342; GB:AE003852; NID:G9657359; PIDN:AAF9590  
 A; Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C; Genetics:  
 A; Gene: VC2765  
 A; Map position: 1  
 C; Superfamily: H+-transporting ATP synthase gamma chain

Query Match 69.9%; Score 998.5; DB 2; Length 288;  
 Best Local Similarity 65.1%; Pred. No. 3.7e-58;  
 Matches 188; Conservative 54; Mismatches 46; Indels 1; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKWKTKQERMSSSRPYSETIRNVISHVSKAT 60  
 DB 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKWKTKQERMSSSRPYSETIRNVISHVSKAT 60

QY 61 IYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTLNEMKEKDVSVQSLIGS 120  
 DB 61 LEYRHPYLEERAKRGVYIIITDRGLCGGLNINLFKKAITDMOTWKEGAQIELAIGS 120

QY 121 KSINFFQSLGKILITQDSIGDTPSVQGLIGSVNSMIDAYKKGEVDVYLVYKFNITMS 180  
 DB 121 KATAFFNNSGAKVAQVSGLDGSLDELIGSVGMUKYKDGELRLYLVFNQFNTMV 180

QY 181 QKPVLEKLIPLDELNDLGERKQWVDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
 DB 181 QKPVLEKLIPLDELNDLGERKQWVDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240

QY 241 EQARWVAMKATDNAGNLNELQVYNKARQASITNELNEIVAGAAAI 289  
 DB 240 EQARWVAMKATDNAGNLNELQVYNKARQASITNELNEIVAGAAAV 288

RESULT 3  
 PHECG  
 H+-transporting two-sector ATPase (EC 3.6.3.14) gamma chain - *Escherichia coli* (strain K-12)  
 C; Species: *Escherichia coli*  
 C; Date: 02-Apr-1982 #sequence, revision 15-Oct-1982 #text\_change 09-Jul-2004  
 C; Accession: A01038; I41276; A01039; I55328; A53160; F65176  
 R; Saraste, M.; Gay, N.J.; Eberle, A.; Runswick, M.J.; Walker, J.E.  
 Nucleic Acids Res. 9, 5287-5296, 1981  
 A; Title: The *atp* operon: nucleotide sequence of the genes for the gamma, beta, and epsilon subunits of the F1F0 ATPase of *Escherichia coli* K-12  
 A; Reference number: A93742; MUID:82059507; PMID:6272217  
 A; Accession: A01038  
 A; Molecule type: DNA  
 A; Residues: 1-287 <SAR>  
 A; Cross-references: UNIPROT:P00837; GB:V00267; NID:G41036; PIDN:CAA23526.1; PID:G41037  
 R; Kanazawa, H.; Futai, M.  
 Ann. N. Y. Acad. Sci. 402, 45-64, 1982  
 A; Title: Structure and function of H+-ATPase: What we have learned from *Escherichia coli*  
 A; Reference number: I41271; MUID:83176724; PMID:6303339  
 A; Accession: I41276  
 A; Status: preliminary; translated from GB/EMBL/DBJ  
 A; Molecule type: DNA  
 A; Residues: 1-287 <RES>  
 R; Kanazawa, H.; Kayano, T.; Mabuchi, K.; Futai, M.  
 Biochem. Biophys. Res. Commun. 103, 604-612, 1981  
 A; Title: Nucleotide sequence of the genes coding for alpha, beta and gamma subunits of the F1F0 ATPase of *Escherichia coli* K-12  
 A; Reference number: A90103; MUID:82134798; PMID:6277310  
 A; Accession: A01039  
 A; Molecule type: DNA  
 A; Residues: 1-37, 'AWRPVLMQKPAQ', 52-82, 'I', 84-92, 'T', 94-111, 'HSTSGAAR', 121, 'D', 123-127  
 A; Cross-references: GB:V00312; NID:G42282; PIDN:CAA33597.1; PID:G42284

R; Iwamoto, A.; Miki, J.; Maeda, M.; Futai, M.  
 J. Biol. Chem. 265, 5043-5048, 1990  
 A; Title: H+-ATPase gamma-subunit of *Escherichia coli*: Role of the conserved carboxyl-terminus in the assembly of the F1F0 ATPase  
 A; Reference number: I55328; MUID:90202983; PMID:2138624  
 A; Accession: I55328  
 A; Status: preliminary; translated from GB/EMBL/DBJ  
 A; Molecule type: DNA  
 A; Residues: 261-287 <RE2>  
 A; Cross-references: GB:M34095; NID:G148146; PIDN:AAA24742.1; PID:G148147  
 R; Tang, C.; Wilkens, S.; Capaldi, R.A.  
 J. Biol. Chem. 269, 4467-4472, 1994  
 A; Title: Structure of the gamma subunit of *Escherichia coli* F1-ATPase probed in trypsin digestion  
 A; Reference number: A53160; MUID:94140880; PMID:7508444  
 A; Accession: A53160  
 A; Status: preliminary  
 A; Molecule type: protein  
 A; Residues: 2-872-81; 82, 204-208; 214-220 <TAN>  
 R; Blatter, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C. A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A; Title: The complete genome sequence of *Escherichia coli* K-12.  
 A; Reference number: A64720; MUID:97426617; PMID:9278503  
 A; Accession: F65176  
 A; Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A; Molecule type: DNA  
 A; Residues: 1-287 <BLAT>  
 A; Cross-references: GB:AE000450; GB:U00096; NID:G1790166; PIDN:AACT6756.1; PID:G1790171  
 A; Experimental source: strain K-12, substrain MG1655  
 C; Genetics:  
 A; Gene: *atpG*; uncG  
 A; Map position: 84 min  
 C; Superfamily: H+-transporting ATP synthase gamma chain  
 C; Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match 69.3%; Score 991; DB 1; Length 287;  
 Best Local Similarity 66.1%; Pred. No. 1.1e-57;  
 Matches 191; Conservative 47; Mismatches 49; Indels 2; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKWKTKQERMSSSRPYSETIRNVISHVSKAT 60  
 DB 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKWKTKQERMSSSRPYSETIRNVISHVSKAT 60

QY 61 IYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTLNEMKEKDVSVQSLIGS 120  
 DB 61 LEYRHPYLEERAKRGVYIIITDRGLCGGLNINLFKKAITDMOTWKEGAQIELAIGS 120

QY 121 KSINFFQSLGKILITQDSIGDTPSVQGLIGSVNSMIDAYKKGEVDVYLVYKFNITMS 180  
 DB 121 KGVSPFNSVGGNVVAQVGTGMGNPSLSLIGPVKVMQLQAYDEGLDKLYIVSNKFNITMS 180

QY 181 QKPVLEKLIPLDELNDLGERKQWVDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
 DB 181 QVPTISQLPLPASDDDL--KHKSWDYLYEPDPKALLDILLRIVSVQYQGVVENLAS 238

QY 241 EQARWVAMKATDNAGNLNELQVYNKARQASITNELNEIVAGAAAI 289  
 DB 239 EQARWVAMKATDNAGNLNELQVYNKARQASITNELNEIVAGAAAV 287

RESULT 4  
 C91213  
 membrane-bound ATP synthase gamma-subunit *atpG* [imported] - *Escherichia coli* (strain O1)  
 C; Species: *Escherichia coli*  
 C; Date: 18-Jul-2001 #sequence, revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C; Accession: C91213  
 R; Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 DNA Res. 8, 11-22, 2001  
 A; Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands  
 A; Reference number: A99629; MUID:21156231; PMID:11258796  
 A; Accession: C91213  
 A; Status: preliminary  
 A; Molecule type: DNA

A;Residues: 1-287 <HAY>  
A;Cross-references: UNIPROT:P00837; GB:BA000007; PIDN:BA038098.1; PID:G13364150; GSPDB:G  
A;Experimental source: strain O157:H7, substrain RIMD 050952  
C;Genetics:  
A;Gene: ECs4675  
C;Superfamily: H+-transporting ATP synthase gamma chain

Query Match 69.3%; Score 991; DB 2; Length 287;  
Best Local Similarity 66.1%; Pred. No. 1.1e-57;  
Matches 191; Conservative 47; Mismatches 49; Indels 2; Gaps 1;  
QY 1 MAGAKEIRTKIASVKSTOKITKAMENVAASVKRKTQERMSRRPYSETIRNVISHVSKAT 60  
DB 1 MAGAKEIRTKIASVKSTOKITKAMENVAASVKRKTQERMSRRPYSETIRNVISHVSKAT 60  
QY 61 IGYKHPFLVDREVKVGMIVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQSLTIGS 120  
DB 61 LEYKHPYLEDRVKRGVYLVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQSLTIGS 120  
QY 121 KSNFFQSLGKILTDQSGIGTTPSVEQLIGSVNSMIDAYKGEVDVYVLYVYVNFINTMS 180  
DB 121 KGVSPFNSVGGNVAVQVTGMGNPSLSLIGPVKVMLOQAYDEGLDKLYIVSNKFINMS 180  
QY 181 QKPVLEKILPLPELDNDELGERKQVNDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
DB 181 QVPTISQLLEPLPASDDDL--KHKSMDYLYEPDPKALLDTLLRRYVESQVYQGVVENLAS 238  
QY 241 EQAARVMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289  
DB 239 EQAARVMVAMKAATDNAGSLIKELQVYNKARQASITQELTEIVSGAAV 287

RESULT 5  
D86059  
membrane-bound ATP synthase gamma-subunit AtpG [imported] - Escherichia coli (strain O157:H7)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: D86059  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: D86059  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-287 <STO>  
A;Cross-references: UNIPROT:P00837; GB:AE005174; NID:G12518590; PIDN:AAG58936.1; GSPDB:G  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: atpG  
C;Superfamily: H+-transporting ATP synthase gamma chain

Query Match 69.3%; Score 991; DB 2; Length 287;  
Best Local Similarity 66.1%; Pred. No. 1.1e-57;  
Matches 191; Conservative 47; Mismatches 49; Indels 2; Gaps 1;  
QY 1 MAGAKEIRTKIASVKSTOKITKAMENVAASVKRKTQERMSRRPYSETIRNVISHVSKAT 60  
DB 1 MAGAKEIRTKIASVKSTOKITKAMENVAASVKRKTQERMSRRPYSETIRNVISHVSKAT 60  
QY 61 IGYKHPFLVDREVKVGMIVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQSLTIGS 120  
DB 61 LEYKHPYLEDRVKRGVYLVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQSLTIGS 120  
QY 121 KSNFFQSLGKILTDQSGIGTTPSVEQLIGSVNSMIDAYKGEVDVYVLYVYVNFINTMS 180  
DB 121 KGVSPFNSVGGNVAVQVTGMGNPSLSLIGPVKVMLOQAYDEGLDKLYIVSNKFINMS 180  
QY 181 QKPVLEKILPLPELDNDELGERKQVNDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
DB 181 QVPTISQLLEPLPASDDDL--KHKSMDYLYEPDPKALLDTLLRRYVESQVYQGVVENLAS 238

QY 241 EQAARVMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289  
DB 239 EQAARVMVAMKAATDNAGSLIKELQVYNKARQASITQELTEIVSGAAV 287  
RESULT 6  
AF0500  
H+-transporting two-sector ATPase (EC 3.6.3.14) gamma chain [imported] - Yersinia pestis  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AF0500  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AF0500  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-287 <KUR>  
A;Cross-references: UNIPROT:Q829S5; GB:AL590842; PIDN:CAC93571.1; PID:G15982011; GSPDB:G  
C;Genetics:  
A;Gene: atpG  
C;Superfamily: H+-transporting ATP synthase gamma chain  
C;Keywords: hydrolase

Query Match 69.2%; Score 989; DB 2; Length 287;  
Best Local Similarity 67.1%; Pred. No. 1.5e-57;  
Matches 194; Conservative 45; Mismatches 47; Indels 2; Gaps 1;  
QY 1 MAGAKEIRTKIASVKSTOKITKAMENVAASVKRKTQERMSRRPYSETIRNVISHVSKAT 60  
DB 1 MAGAKEIRTKIASVKSTOKITKAMENVAASVKRKTQERMSRRPYSETIRNVISHVSKAT 60  
QY 61 IGYKHPFLVDREVKVGMIVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQSLTIGS 120  
DB 61 LEYKHPYLEDRVKRGVYLVSTDRGLCGGLNNVLFKTVLNMKEWKEKDVSVQSLTIGS 120  
QY 121 KSNFFQSLGKILTDQSGIGTTPSVEQLIGSVNSMIDAYKGEVDVYVLYVYVNFINTMS 180  
DB 121 KAASFPQSGVGKIVAVQVTGMGNPSLSLIGPVKVMLOQAYDEGLDKLYIVNNKFINMS 180  
QY 181 QKPVLEKILPLPELDNDELGERKQVNDYIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
DB 181 QEPRIWQLPLPPAEDGEL--KKSWDYLYEPDPKALLDTLLRRYVESQVYQGVVENLAS 238  
QY 241 EQAARVMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289  
DB 239 EQAARVMVAMKAATDNAGSLIKELQVYNKARQASITQELTEIVSGASAV 287

RESULT 7  
AE0954  
ATP synthase gamma chain [imported] - Salmonella enterica subsp. enterica serovar Typhi  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AE0954  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Conner, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moul, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AE0954  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-287 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD03129.1; PID:G16504765; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY3912

C/Superfamily: H<sup>+</sup>-transporting ATP synthase gamma chain

Query Match	68.4%;	Score 978;	DB 2;	Length 287;
Best Local Similarity	64.7%;	Pred. No. 8e-57;		
Matches 187;	Conservative 51;	Mismatches 49;	Indels 2;	Gaps 1
QY	1	MAGAKEIRTKIASVKVSTQKTIKAMEMVAASQWRKTOBMRSSSRPYPGETIRNWIHVSXKAT	60	
Db	1	MAGAKEIRSKIASVQNTQKTIKAMEMVAASQWRKSDQDRAASPYAETWRKVIHGLANGN	60	
QY	61	LYGKHPPFLVDSEVKVGMIVVSTDSGICGGINVNILFKTVLNEKWEKWKDVSQVLSLGS	120	
Db	61	LEYKHPYLEERADVARKVGLVYVSTDCGICGGINLNILFKQLADWKASDSDXGVCELANTGS	120	
QY	121	KSINFFQSLGIIKIITQDSGIGDTPSVEBQLGYSVNSMIDAYKGEVDVYVLYVNNKKTINTWS	180	
Db	121	KGVSFNFSVSGVNVVAQVTGMGDNPSLSBELIGPVKWLQAYDEGLDKLIYSNKKKTINTWS	180	
QY	181	OKPVLKELIPIPELDNDBELGERKQWVDYIYEPDAKVLNLLVRYLESQVYQAAVENVLAS	240	
Db	181	QVPIITQLLPAPASEDDDL--KRTAWDYLIEPDPKALLDTLLRRYVSSQVYGGVVENVLAS	238	
QY	241	EOAARMVAMKAATNAGNLINELQLVYNKARQASITNELNEIVAGAAAI	289	
Db	239	EOAARMVAMKAATDNGSSLKEQLQVYNKARQASITQELITEIVSGAAAV	287	

RESULT 8  
S06081  
H-H-transporting two-sector ATPase (EC 3.6.3.14) gamma chain - Vibrio alginolyticus  
C:Species: Vibrio alginolyticus  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: S06081  
R:Krumholz, L.R.; Esser, U.; Simoni, R.D.  
Nucleic Acids Res. 17, 7993-7994, 1989  
A:A:Title: Nucleotide sequence of the unc operon of Vibrio alginolyticus.  
A:A:Reference number: S06075; MUID:90016889; PMID:2529481  
A:Accession: S06081  
A:A:Status: translation not shown  
A:A:Molecule type: DNA  
A:A:Residues: 1-288 KRU>  
A:A:Cross-references: UNIPROT:P12990; EMBL:X16050; NID:G48331; PID:CAA34180.1; PID:G48338  
C:Genetics:  
A:Gene: uncG  
C:Superfamily: H-transporting ATP synthase gamma chain  
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match	68.3%;	Score	976.5;	DB 1;	Length	288;			
Best Local Similarity	63.3%;	Pred. No.	1e-56;						
Matches	183;	Conservative	56;	Mismatches	49;	Indels	1;	Gaps	1;
2y	1	MAGAKIIRTKIAGSVKQTQKITKAMENVAASMKRKQERMSRSSRPVSEIRTNVISHVSKAT	60						
ob	1	MAGAKEIRNKIGSVKSTQKITKAMENVAASMKRSQDAMEASRPYAEITMKVIGHGVANAN	60						
2y	61	IGYKHPFLDREVKKVGMIWSTDRGLCGGLNNVLFKTVLNMKEWKKBQVSVQLSLGS	120						
ob	61	LEYRHPFLBEREAKRVGYIIVSTDRLGCGGLNNVFKKAVTDMQTRWKGAEIELAVVGS	120						
2y	121	KSINFFQSGLKILTDQSGIGTTPSVEQIGLGSVNSMIDAYKKGBVDVVVLYVYKFKINTMS	180						
ob	121	KATAFFKHGSAKVAAGSVSLGDNPSLEDLIGSVGMLKKYDEGELDRLYVVFNFKVNTMV	180						
2y	181	QKPVLEKLIPLPELNDDELGEKQKQWDVIYEPDQAVLLDNLVLYLESQVYQAQAVENLAS	240						
ob	181	QQTIQDILLPKPSUSEEM-QREHSWDVIYEPPEKPLIDTLVLYVESQVYQGVVENLAC	239						
2y	241	EQASRVAKKATDNAGNLINELQLVYVKARQAASITNELNEIVAGAAA	289						
b	240	EQAARMIAKKAATDNATNLIDLELVYVKARQAAITQELSEIVGGAAV	288						

## RESULT 9

D82952  
ATP synthase gamma chain PA5555 [imported] - *Pseudomonas aeruginosa* (strain PAO1)  
C:Species: *Pseudomonas aeruginosa*  
C:date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: D82952  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic path  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: D82952  
A:status: preliminary  
A:status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-286 <STO>  
A:cross-references: UNIPROT:Q9HTL9; GB:AE004967; GB:AE004091; NID:g9951884; PIDN:AAG089,  
A:Experimental source: strain PAO1  
C:Genetics:  
A:Gene: atpG; PA5555  
C:Superfamily: H+-transporting ATP synthase gamma chain

Query Match	64.8%	Score	926.5	DB 2	Length	286
Best Local Similarity	62.3%	Prod. No.	1.8e-53			
Matches 180	Conservative	50	Mismatches	56	Indels	3
					Gaps	1
QY	1	MAGAKEIRTIKIASVKSTQKITKAMEMVAASKVRKQTCERMSSRPYSEIRNVISHVSKAT	60			
DB	1	MAGAKEIRSKIASIKSTQKITNAMKEMVAVSKRKQKQMAAGRPFAEIRQVIGHLANAN	60			
QY	61	IGYKHPFLVDREVKYGMIVVSTDRLCGGLNVNFKTVLNMKEWKQKDVSVLSIGS	120			
DB	61	PEYRHPFMVERKVRGVYIVVSDRLCGGLNINLFKSLVKDMGSRQGAEDLCVIGS	120			
QY	121	KSINFFOSLGKILTDOSGIGTPPSVEQULIGSWNSKIDAYKKGEVDVVYLVNKFINMTS	180			
DB	121	KGASFRRFSGCNVAAISHLGEPSINDLIGSKVMDLAYLEGRIDLFFVYSNKFVNTYT	180			
QY	181	QKPVLEKLIPLPBLDNDELGERKQVWDYIYEPDAKVLNLLVYLSOVYCAAVENLAS	240			
DB	181	QKPTVEQLIFLVADDDQDEL---KHHWDYIYEPDAKSLLDGLLVRFVSQVYQAVENNAC	237			
QY	241	EQARBMVAMKAATDNAGNLINELQLVYTKARQAOSTITNELNEIVAGAAAI	289			
DB	238	EQARMTAMKATDNAGELISLQIYTKARQAATQEISIVGGAANV	286			

```

RESULT 10
G84930
H+H+-transporting two-sector ATPase (BC 3.6.3.14) gamma chain [imported] - Buchnera sp. (
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 03-Jun-2002
C:Accession: G84930
R:R.Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Tshikawa, H.
N:Nature 407, 81-85, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. ?
A:Reference number: A84930; PMID:20445173; PMID:10993077
A:Accession: G84930
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <STO>
A:Cross-References: GB:AP000398; GSPDB:GM00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: atpG; BU007
H+:Superfamily: H+-transporting ATP synthase gamma chain
Keywords: Hydrolase

```

```
Query Match          56.5%; Score 808; DB 2; Length 290;
Best Local Similarity 52.6%; Pred. No. 1e-45;
Matches 152; Conservative 60; Mismatches 75; Indels 2; Gaps 1;
```

Db 1 MTSTKIKNKIVSVNTKTKITKAMEMVAVSKVRKTEERMRSGRPYSDIIRKVIDHTVQGN 60

QY 61 IGYKHPLVDREVKKYGMVSTDRGLCGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 120

Db 61 LEYKHSYLERKNTKIRNIGMIIISTDRGLCGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 120

QY 121 KSNFFQSLGKILTOPDSGIGTPSPVEQVQLIGSVNSMIDAYKKGEVDVVYLVYVNFINTM 180

Db 121 KSLSVFKLGSNILAKATNLGENPKLEELNSVGILOEQYCKRIDKFIAYNFKHNM 180

QY 181 QKPVLEKLPPLDELNDLGERKQVWDIYEPDAKVLDNLLVYLSQVYQAAVENLAS 240

Db 181 QPTIQLPFFSKQDDASNN--WDVYEPESKLIIDTLFNRYISQVQSIENIAS 238

QY 241 EQAARVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289

Db 239 EHAARMIAMKTATNSGNRIKELQVYNKVRQANITQELNEIVSGASAV 287

RESULT 11

H82715

ATP synthase, gamma chain XF1144 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

C:Accession: H82715

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

Nature 405, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; PMID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82715

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-287 <S>

A:Cross-references: UNIPROT:Q9PB84; GB:AE003950; GB:AE003849; NID:G9106104; PIDN:AAF8395

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, H.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Doroty, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuramae, E.E.; Laigh chado, M.A.; Madeira, M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1144

C:Superfamily: H+-transporting ATP synthase gamma chain

Query Match 55.6%; Score 795; DB 2; Length 287;

Best Local Similarity 52.4%; Pred. No. 7e-45;

Matches 154; Conservative 66; Mismatches 62; Indels 12; Gaps 3;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAVSKVRKTEERMSSSRPYSETIRNVISHVSKAT 60

Db 1 MASGREIKSKIKSVNTKTRVTRALEMVSASIKRKAQEQMKISRPYQAQNKQWTHLAQAN 60

QY 61 IGYKHPLVDREVKKYGMVSTDRGLCGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 119

Db 61 TDYUHPFLIAHKQVRIKGYIVISDRGLAGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 120

QY 120 SKSNFFQSLGKILTOPDSGIGTPSPVEQVQLIGSVNSMIDAYKKGEVDVVYLVYVNFINTM 179

Db 121 QKASVFFERIKNVLGVTHTLGTTPRLQLIGIKVLMDAYTEEKLDRLVYVNFINTM 180

QY 180 SQKPVLEKLPPLDELNDLGERKQVWDIYEPDAKVLDNLLVYLSQVYQAAVENLAS 235

Db 181 VQKASFDQLLPL-----LAAKDKVAHHDWDVLYEPDAATVLEHMRRYTESLYQAWL 233

QY 236 ENLASEQARVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289

Db 234 ENIASEHAARVAMKAASDNANKLIGTLQVYNKARQAAITQEISEIVGGAAG 287

RESULT 12

H81970

H+-transporting two-sector ATPase (EC 3.6.3.14) gamma chain NMA0518 [imported] - Neisser

C:Species: Neisseria meningitidis

C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004

C:Accession: H81970

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel ; Holroyd, S.; Jagsels, K.; Leather, S.; Moutle, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; PMID:20222556; PMID:10761919

A:Accession: H81970

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-291 <PAR>

A:Cross-references: UNIPROT:Q9JW71; GB:AL162753; GB:AL157959; NID:G7379120; PIDN:CA88381

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: atpG; NMA0518

C:Superfamily: H+-transporting ATP synthase gamma chain

C:Keywords: hydrolase

Query Match 52.8%; Score 754; DB 2; Length 291;

Best Local Similarity 51.2%; Pred. No. 3.4e-42;

Matches 149; Conservative 64; Mismatches 76; Indels 2; Gaps 2;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAVSKVRKTEERMSSSRPYSETIRNVISHVSKAT 60

Db 1 MAVGKEILTKIRSVNTQKITKAWQVSTSKMEKTOERMELAPPYAEKVRWVWVSHLAQTN 60

QY 61 IGYKHPLVDREVKKYGMVSTDRGLCGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 119

Db 61 TDHGIPLESHEIRRVGFIITLSDKGLCGLNVLNFKTVLNMKEWKEKDVSVQSLIGS 120

QY 120 SKSNFFQSLGKILTOPDSGIGTPSPVEQVQLIGSVNSMIDAYKKGEVDVVYLVYVNFINTM 179

Db 121 SKGLMACQSLGNVAVSAYNLGTPKWEMLGLPLTLFQYKKEHIDRLHLYVSGFVNTM 180

QY 180 SQKPVLEKLPPLDELNDLGERKQVWDIYEPDAKVLDNLLVYLSQVYQAAVENLAS 238

Db 181 RQEPREMEVLLPIGENVIGDSAPKSPFSWRYEPTALAVLYLVRYLESVYVYQALSDNM 240

QY 239 ASEQAARVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289

Db 241 ASEQAARVAMKAATDNAGNAIKELVYNKARQAAITTEISEIVAGAAAV 291

RESULT 13

H81024

ATP synthase F1, gamma chain NMB1935 [imported] - Neisseria meningitidis (strain MC58 se

C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004

C:Accession: H81024

R:Tettelein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A Hickey, E.K.; Hait, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M. Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; PMID:20175755; PMID:10710307

A:Accession: H81024

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-291 <TET>

A:Cross-references: UNIPROT:Q9JX01; GB:AE002542; GB:AE002098; NID:G7227188; PIDN:AAF4226.

A:Experimental source: serogroup B, strain MC58

C:Genetics:

ATP synthase gamma subunit atpG [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C:Accession: C84119  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hise,  
Nucleic Acids Res. 28, 4317-4331, 2000  
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: AB3650; MUID:20512582; PMID:11058132  
A:Accession: C84119  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-285 <STO>  
A:Cross-references: UNIPROT:Q9K6H4; GB:AB001519; GB:BA000004; NID:g10176109; PIDN:BAB07.2  
A:Experimental source: strain C-125  
C:Genetics:  
C:A Gene: atpG  
C:Superfamily: H+-transporting ATP synthase gamma chain

Query Match 37.5%; Score 536.5; DB 2; Length 285;  
Best Local Similarity 40.0%; Pred. No. 5.2e+28;  
Matches 116; Conservative 63; Mismatches 104; Indels 7; Gaps 3;

QY 1 MAGAKEIRTKIASVSKTOKITKAMEMVAASKVKRKTOERMSSRPYSSETIRNVISHVSKAT 60  
Db :  
1 MASLRDIKQRINSTRKKTKKITKAMEMVAASKNRSGEKAQSFLPYTDKIREVVASPAAD 60  
QY 61 IGYKHPLVDREKVKGVMIVSTDRLGCGLNVNFKTVLNM-KEWKEKDVSVOLSLIG 119  
Db TVDSHPMLSERPVKKTGYIVITSDRLAGAYNSNLIRGLLYTINKRHKSKD-EYGIFAIG 119  
QY 120 SKSINFPQSLGIKLTODSGIGDTSPVEGLIGSVNSMIDAYKKGEVDVVLVYNKFINTM 179  
Db RTGRDLLKKRQPLISEMTGLSDQTFFNDIKIAQTVDMPFADEVFDELIYIWNHFVSP 179  
QY 180 SKQPVLKLIPIPELDNDDELGERKQWPTYIEPDAKVLLDNLLRYLESQVYCAAVENLA 239  
Db 180 KDQVTEKKVLPITDLSDTKVST-----TVEYEPNEQVILEALLPOVASLVYGALLDAKA 234  
QY 240 SPQAARMVAMKAATNAGNLNELOLVYNKARQASITNELNEIVAGAAAI 289  
Db 235 SEFAARMVAMSAAATNATNDELTLTSYNRAQAAITOETIEIVGGAAL 284

Search completed: November 9, 2004, 09:53:37  
Job time : 17 secs

A:Gene: NMB1935  
C:Superfamily: H<sup>+</sup>-transporting ATP synthase gamma chain

Query Match	52.7%;	Score 753;	DB 2;	Length 291;
Best Local Similarity	51.2%;	Pred. No. 4e-42;		
Matches 149;	Conservative 63;	Mismatches 77;	Indels 2;	Gaps 2;

QY 1 MAGAKRIKIAVSKTQKTIKAMVVAASKMKTQERMSSSSPYSETIRNVISHVSKAT 60  
Db 1 MAYGKEILTIRSVQNTQKTIKAMQVYSTSKMKTQERMRLARPYAEKVWVNSHLAQTN 60  
QY 61 IGYKHPEFLVD-REVKKVGMIVWSTDGKGLGNVNLFKTVLNMKWKKKDVSVQLSLIG 119  
Db 61 TDHGIPLLASHREIRRVGFILITSDKGLGNANVVKFLAQVQRYRQGTIEEIVCFG 120  
QY 120 SKSINFFQSLGIKILITQDSIGITPTPSVEOLIGSVNSMIDAYKKGEVDVYLVYNKFINTM 179  
Db 121 SKGLMACQSIGLNVVASAVNLGTPKQEMLLGLPTLSELFQRYEKHEIDRIHLVYSGFVNTM 180  
QY 180 SOKPVLKELIPLPE-LDNDLGBGRKQWVDYVBPDAKVLDDNLLVRLYSQVYQVAENVL 238  
Db 181 REPRFNEVLPTGSEVIGVSDAPKSPSEWRYEPTALAVLEVLYRRLYSVVIQALSDNN 240  
QY 239 ASEQAARVMYAMKAATDNAGNLINELQVLYNKAFQASITNMLNEIVAGAAAI 289  
Db 241 ASEQAARVMYAMKAATDNAGNAIKELRLVYNKSRQAATTSELSEIVAGAAAV 291

RESULT 14  
G31482  
H<sup>+</sup>-transporting two-sector ATPase (EC 3.6.3.14) gamma chain - Bacillus megaterium  
C:Species: Bacillus megaterium  
C:Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 09-Jul-2004  
C:Accession: G31482  
R:Brusilow, W.S.A.; Scarpetta, M.A.; Hawthorne, C.A.; Clark, W.P.  
J. Biol. Chem. 264: 1528-1533, 1989  
A:Title: Organization and sequence of the genes coding for the proton-translocating ATPase  
A:Reference number: A31482; MUID:89109162; PMID:2521483  
A:Accession: G31482  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-285 <BRU>  
A:Cross-references: UNIPROT:P20602; GB:M20255; GB:J04455; GB:M18352; GB:M23924; NID:g142  
C:Superfamily: H<sup>+</sup>-transporting ATP synthase gamma chain  
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

[illegible]

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2004, 09:47:28 ; Search time 193 Seconds  
(without alignments)  
861.571 Million cell updates/sec

Title: US-09-545-199F-4

Perfect score: 1429

Sequence: 1 MAGAKIRTKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1429	100.0	289	1 ATPG_PASMU	Q916b6 pasteurella
2	1173	82.1	289	1 ATPG_HAEIN	P43716 haemophilus
3	1121.5	78.5	288	2 Q7VPP1	Q7vpp1 haemophilus
4	1004	70.3	287	2 Q7NA93	Q7na93 photorhabdu
5	998.5	69.9	288	2 Q9KNH4	Q9knh4 vibrio chol
6	993	69.5	287	2 Q6CVJ4	Q6cvj4 erwiania car
7	991	69.3	287	1 ATPG_ECOLI	P00837 escherichia
8	991	69.3	287	2 Q9PFI4	Q9pfi4 salmonella
9	989	69.2	287	2 Q8Z9S5	Q8z9s5 yersinia pe
10	989	69.2	287	2 A8S64168	A8s64168 yersinia
11	988	69.1	287	2 Q8ZKX8	Q8zkx8 salmonella
12	982.5	68.8	286	2 Q8E8B9	Q8eb9 shewanella
13	980.5	68.6	288	2 Q7MGH9	Q7mgh9 vibrio vuln
14	980.5	68.6	288	2 Q8DDG9	Q8ddg9 vibrio vuln
15	978	68.4	287	2 Q822Q5	Q822q5 salmonella
16	976.5	68.3	288	1 ATPG_VIBAL	P12990 vibrio algi
17	970.5	67.9	288	2 Q87KA7	Q87ka7 vibrio para
18	947.5	66.3	288	2 Q6L7G7	Q6l7g7 photobacter
19	947.5	66.3	288	2 CAG21861	Cag21861 photobact
20	931.5	65.2	287	2 Q8VW78	Q8vw78 colwellia m
21	926.5	64.8	286	2 Q8HT19	Q8ht19 pseudomonas
22	915	64.0	291	2 Q4KZ27	Q4kz27 photobacter
23	915	64.0	291	2 CAG22008	Cag22008 photobact
24	896.5	62.7	286	2 Q88BX3	Q88bx3 pseudomonas
25	889	62.2	289	2 Q8FFK1	Q8ffk1 acinetobact
26	885.5	62.0	286	2 Q87TT3	Q87tt3 pseudomonas
27	885.5	62.0	294	2 Q82XP9	Q82xp9 nitrosomona
28	885	61.9	291	2 Q83AF5	Q83af5 ralstonia s
29	847.5	59.3	289	2 Q8QA76	Q8qa76 coxiella bu
30	838	58.6	287	2 Q8FC26	Q8fc26 xanthomonas
31	830	58.1	287	2 Q8P6G6	Q8p6g6 xanthomonas

32	808	56.5	290	1 ATPG_BUCAI	P57123 buchnera ap
33	805	56.3	301	2 Q7VU45	Q7vu45 bordetella
34	804	56.3	301	2 Q7WEM8	Q7wem8 bordetella
35	802	56.1	301	2 Q7WJA9	Q7wja9 bordetella
36	799.5	55.9	298	1 ATPG_THIFE	P41169 thiobacillu
37	795	55.6	287	2 Q9PE84	Q9pe84 xylella fas
38	791	55.4	287	2 Q87E89	Q87e89 xylella fas
39	790.5	55.3	291	1 ATPG_BUCAP	C51873 buchnera ap
40	783.5	54.8	288	2 Q7VQV7	Q7qv7 candidatus
41	754	52.8	291	2 Q9NW71	Q9nw71 neisseria m
42	753	52.7	291	2 Q9UXQ1	Q9uxq1 neisseria m
43	751.5	52.6	275	2 Q7P096	Q7p096 chromobacte
44	747	52.3	290	2 Q8RQ80	Q8rq80 buchnera ap
45	693.5	48.5	287	2 Q8D3J4	Q8dj34 wiggleswort

## ALIGNMENTS

### RESULT 1

ATPG\_PASMU STANDARD; PRT; 289 AA.

AC Q9L6B6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 03-JUL-2004 (Rel. 44, Last annotation update)  
DE ATP synthase gamma chain (EC 3.6.3.14).  
GN Name=atpg; OrderedLocusNames=PM1493;  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fuller T.E., Kennedy M.J., Lowery D.E.;  
RT "Identification of Pasteurella multocida virulence genes in a  
septicemic mouse model using signature-tagged mutagenesis."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Pm70;  
RX MEDLINE=21145866; PubMed=11248100; DOI=10.1073/pnas.051634598;  
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida Pm70."  
Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
CC 1- FUNCTION: Produces ATP from ADP in the presence of a proton  
gradient across the membrane. The gamma chain is believed to be  
important in regulating ATPase activity and the flow of protons  
through the CF(0) complex.  
CC 1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +  
H(+) (Out).  
CC 1- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic  
core - and CF(0) - the membrane proton channel. CF(1) has five  
subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)  
has three main subunits: a, b and c.  
CC 1- SIMILARITY: Belongs to the ATPase gamma chain family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@sib-sib.ch).  
CC -----  
CC EMBL; AF237922; AAF68408.1; -  
CC EMBL; AE006186; AAK03577.1; -  
CC HSPF; P00837; 1FS0.  
CC InterPro; IPR000131; ATPase\_gamma.  
CC Pfam; PF00231; ATP-synt; 1.  
CC PRINTS; PR00126; ATPASEGAMMA.  
CC TIGRFAMs; TIGR01146; ATPsyn\_F1gamma; 1.  
CC PROSITE; PS00153; ATPASE\_GAMMA; 1.

KW ATP synthesis; CF(1); Complete proteome; Hydrogen ion transport;  
 KW Hydrolyase. 289 AA; 32095 MW; 1E1B862B4E9A9F70 CRC64;  
 SQ SEQUENCE

Query Match 100.0%; Score 1429; DB 1; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-86;  
 Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKRKTOERMSSSRPYSETIRNVISHVSKAT 60  
 DB 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKRKTOERMSSSRPYSETIRNVISHVSKAT 60  
 QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLFKTVLNEMKEWKEDVSVQLSLIGS 120  
 DB 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLFKTVLNEMKEWKEDVSVQLSLIGS 120  
 QY 121 KSNFFQSGIGIKILTQDSGIGDTPSPVEQLIGSVNSMIDAYKGEVDVYLVYVNFINTMS 180  
 DB 121 KSNFFQSGIGIKILTQDSGIGDTPSPVEQLIGSVNSMIDAYKGEVDVYLVYVNFINTMS 180  
 QY 181 QKPVLEKLIPLPELDNDELGERKQVWDIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
 DB 181 QKPVLEKLIPLPELDNDELGERKQVWDIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
 QY 241 EQAARVMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289  
 DB 241 EQAARVMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289

RESULT 2  
 ATPG\_HABIN STANDARD; PRT; 289 AA.  
 AC P43776;  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE ATP synthase gamma chain (EC 3.6.3.14).  
 GN Name=atpG; OrderedLocusNames=HI0480;  
 OS Haemophilus influenzae  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Keriavay A.R., Ehalt C.O., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G.G., FitzHugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Frithman J.L., Fuhrmann J.L., Geiselman N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 RC -!- FUNCTION: Produces ATP from ADP in the presence of a proton  
 gradient across the membrane. The gamma chain is believed to be  
 important in regulating ATPase activity and the flow of protons  
 through the CF(0) complex.  
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +  
 H(+) (Out).  
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic  
 core - and CF(0) - the membrane proton channel. CF(1) has five  
 subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)  
 has three main subunits: a, b and c.  
 CC -!- SIMILARITY: Belongs to the ATPase gamma chain family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U32730; AAC22138.1; -.  
 DR F1R; E64071; E64071.  
 DR TIGR; H10480; -.  
 DR InterPro; IPR000131; ATPase\_gamma.  
 DR Pfam; PF00231; ATP-synt. 1.  
 DR PRINTS; PR0126; ATPASEGAMMA.  
 DR TIGRFAMs; TIGR01146; ATPsyn\_Figamma; 1.  
 DR PROSITE; PS00153; ATPASE\_GAMMA; 1.  
 KW ATP synthesis; CF(1); Complete proteome; Hydrogen ion transport;  
 KW Hydrolyase.  
 SQ SEQUENCE 289 AA; 32069 MW; 622CBA682F37FD00 CRC64;  
 Query Match 82.1%; Score 1173; DB 1; Length 289;  
 Best Local Similarity 76.5%; Pred. No. 5.1e-69;  
 Matches 221; Conservative 42; Mismatches 26; Indels 0; Gaps 0;

QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKRKTOERMSSSRPYSETIRNVISHVSKAT 60  
 DB 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKRKTOERMSSSRPYSETIRNVISHVSKAS 60  
 QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLFKTVLNEMKEWKEDVSVQLSLIGS 120  
 DB 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLFKTVLNEMKEWKEDVSVQLSLIGS 120  
 QY 121 KSNFFQSGIGIKILTQDSGIGDTPSPVEQLIGSVNSMIDAYKGEVDVYLVYVNFINTMS 180  
 DB 121 KSNFFQSGIGIKILTQDSGIGDTPSPVEQLIGSVNSMIDAYKGEVDVYLVYVNFINTMS 180  
 QY 181 QKPVLEKLIPLPELDNDELGERKQVWDIYEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
 DB 181 QKPVVQVLPLPESKDHLENERQOTDYLVEPEPKVLLDLSLLVRYLESQVYQAAVENLAS 240  
 QY 241 EQAARVMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289  
 DB 241 EQAARVMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289

RESULT 3  
 Q7VPP1 PRELIMINARY; PRT; 288 AA.  
 AC Q7VPP1;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE ATP synthase gamma chain.  
 GN Name=atpG; OrderedLocusNames=HD0009;  
 OS Haemophilus ducreyi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=730;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=35000HP / ATCC 700724;  
 RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,  
 RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;  
 RT "The complete genome sequence of Haemophilus ducreyi.";  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE017151; AAP95033.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.  
 DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; IEA.  
 DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. ; IEA.  
 DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.  
 DR InterPro; IPR000131; ATPase\_gamma.  
 DR Pfam; PF00231; ATP-synt. 1.  
 DR TIGRFAMs; TIGR01146; ATPsyn\_Figamma; 1.  
 DR PROSITE; PS00153; ATPASE\_GAMMA; 1.



KW Complete proteome.  
SQ SEQUENCE 288 AA; 31953 MW; 6CC7342402D4C977 CRC64;

Query Match 78.5%; Score 1121.5; DB 2; Length 288;  
Best Local Similarity 73.4%; Pred. No. 1.2e-65;  
Matches 212; Conservative 48; Mismatches 28; Indels 1; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASVKRKTQERMSSSRPYSETIRNVISHVSKAT 60  
DB 1 MAGAKEIRTKIASVQNTQKITKAMEMVAASVKRKTQDRMAASRPYAEITRSVIGELALGN 60

QY 61 IGYKHPFLVDREVKVGMIVVSTDRGLCGGLNVNLFKTVLNEMKEWKEKDVSVOLSLIGS 120  
DB 61 LEYKHPYLERETKRGVGVVSTDRGLCGGLTNLFKLLSEMDKWDKDVQCELALIGS 120

QY 121 KSNFQSLGKILTDGSGIGDTPSVEQLIGSVNMDIAYKKGEVDVYLVYKFNFTMS 180  
DB 121 KATSPFASVGGNVVAQVVTGMGNPFLSELIGVNMILRAYDEGLDKLYVYVTKFNFTMS 180

QY 181 QXPVLEKLIPLPELDNDELGERKQVWDIYEPDAKVLIDNLLVRYLESQVYQAAVENLAS 240  
DB 181 QETITQLPLPAGDDEL--KKSWDYLYEPDPKALLDILRRYVESQVYQGVVENLAS 238

QY 241 EQAARWVAKAATDNAGNINELQLVYKARQASITNELNEIVAGAAAI 289  
DB 239 EQAARWVAKAATDNGSLIKELQLVYKARQASITQELTEIVSGASAV 287

RESULT 5  
Q9KNH4  
ID Q9KNH4 PRELIMINARY; PRT; 288 AA.  
AC Q9KNH4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE ATP synthase F1, gamma subunit.  
GN OrderedLocustNames=VC2765;  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,  
RA Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I.,  
RA Sellers P., McDonald L.A., Utterback T.R., Fleischmann R.D.,  
RA Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,  
RA Mekalanos J.J., Venter J.C., Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
RT cholerae".  
RL Nature 406:477-483 (2000).  
DR EMBL; AB004342; AAF95904.1; -.  
DR PIR; G82036; G82036.  
DR HSSP; P00837; 1FS0.  
DR TIGR; VC2765; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.  
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; IEA.  
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. ; IEA.  
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.  
DR InterPro; IPR000131; ATPase\_gamma.  
DR Pfam; PF00231; ATP-synt; 1.  
DR PRINTS; PR00126; ATPASEGAMMA.  
DR TIGRFAMs; TIGR01146; ATPsyn\_Figamma; 1.  
DR PROSITE; PS00153; ATPASE\_GAMMA; 1.  
DR Complete proteome.  
SQ SEQUENCE 288 AA; 31862 MW; 2CCE2F03B9982DBF CRC64;

Query Match 69.9%; Score 998.5; DB 2; Length 288;  
Best Local Similarity 65.1%; Pred. No. 1.4e-57;  
Matches 188; Conservative 54; Mismatches 46; Indels 1; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASVKRKTQERMSSSRPYSETIRNVISHVSKAT 60  
DB 1 MAGAKEIRTKIGSVKSTQKITKAMEMVAASVKRKTQDMASRPYATIRKVIQGVANAS 60

KW Complete proteome.  
SQ SEQUENCE 288 AA; 31953 MW; 6CC7342402D4C977 CRC64;

Query Match 78.5%; Score 1121.5; DB 2; Length 288;  
Best Local Similarity 73.4%; Pred. No. 1.2e-65;  
Matches 212; Conservative 48; Mismatches 28; Indels 1; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASVKRKTQERMSSSRPYSETIRNVISHVSKAT 60  
DB 1 MAGAKEIRTKIASVQNTQKITKAMEMVAASVKRKTQDRMAASRPYAEITRSVIGELALGN 60

QY 61 IGYKHPFLVDREVKVGMIVVSTDRGLCGGLNVNLFKTVLNEMKEWKEKDVSVOLSLIGS 120  
DB 61 LEYKHPYLERETKRGVGVVSTDRGLCGGLTNLFKLLSEMDKWDKDVQCELALIGS 120

QY 121 KSNFQSLGKILTDGSGIGDTPSVEQLIGSVNMDIAYKKGEVDVYLVYKFNFTMS 180  
DB 121 KATSPFASVGGNVVAQVVTGMGNPFLSELIGVNMILRAYDEGLDKLYVYVTKFNFTMS 180

QY 181 QXPVLEKLIPLPELDNDELGERKQVWDIYEPDAKVLIDNLLVRYLESQVYQAAVENLAS 240  
DB 181 QETITQLPLPAGDDEL--KKSWDYLYEPDPKALLDILRRYVESQVYQGVVENLAS 238

QY 241 EQAARWVAKAATDNAGNINELQLVYKARQASITNELNEIVAGAAAI 289  
DB 239 EQAARWVAKAATDNGSLIKELQLVYKARQASITQELTEIVSGASAV 287

RESULT 4  
Q7NA93  
ID Q7NA93 PRELIMINARY; PRT; 287 AA.  
AC Q7NA93;  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DE ATP synthase gamma chain.  
GN Name=atpG; OrderedLocustNames=plu0041;  
OS Photorhabdus luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=141679;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TT01;  
RX MEDLINE=22557627; PubMed=14528314;  
RA Duchaud E., Rusnlok C., Frangeul L., Buchrieser C., Givaudan A.,  
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,  
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,  
RA Medigue C., Lanois A., Powell K., Siquier P., Vincent R., Wingate V.,  
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,  
RT "The genome sequence of the entomopathogenic bacterium *Photorhabdus*  
RT luminescens".  
RL Nat. Biotechnol. 21:1307-1313 (2003).  
DR EMBL; BX571859; CAE12336.1; -.  
DR PhotoList; plu0041; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.  
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; IEA.  
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. ; IEA.  
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.  
DR InterPro; IPR000131; ATPase\_gamma.  
DR Pfam; PF00231; ATP-synt; 1.  
DR PRINTS; PR00126; ATPASEGAMMA.  
DR TIGRFAMs; TIGR01146; ATPsyn\_Figamma; 1.  
DR PROSITE; PS00153; ATPASE\_GAMMA; 1.  
DR PROSITE; PS00041; HTH\_ARAC\_FAMILY\_1; UNKNOWN\_1.  
DR Complete proteome.  
SQ SEQUENCE 287 AA; 31644 MW; D957DBEF07B7F415 CRC64;

Query Match 70.3%; Score 1004; DB 2; Length 287;  
Best Local Similarity 67.5%; Pred. No. 6.1e-58;  
Matches 195; Conservative 48; Mismatches 44; Indels 2; Gaps 1;

P00837; P00838;  
21-JUL-1986 (Rel. 01, Created)  
21-JUL-1986 (Rel. 01, Last sequence update)  
01-OCT-2004 (Rel. 45, Last annotation update)  
ATP synthase gamma chain (EC 3.6.3.14).  
Name=atpG; Synonyms=uncG, pspC;  
OrderedLocNames=B5733, c4659, z5231, ECs4675, SF3813, S3955;  
Escherichia coli, O6,  
Escherichia coli O157:H7, and  
Shigella flexneri.  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
NCBI\_TaxID=562, 217992, 83334, 623;  
[1] \_TaxID=562, 217992, 83334, 623;  
SEQUENCE FROM N.A.  
SPECIES=E.coli;  
MEDLINE=85121806; PubMed=6395859;  
Walker J.E., Gay N.J., Saraste M., Eberle A.N.;  
"DNA sequence of the *Escherichia coli* unc operon. Completion of  
the sequence of a 17 kilobase segment containing *asnA*, *oriC*, *unc*, *glms*  
and *phoS*.";  
Biochem. J. 224:739-815 (1984).  
[2]  
SEQUENCE FROM N.A.  
SPECIES=E.coli;  
MEDLINE=82059507; PubMed=6272217;  
Saraste M., Gay N.J., Eberle A., Runswick M.J., Walker J.E.;  
"The *atp* operon: nucleotide sequence of the genes for the gamma, beta,  
and epsilon subunits of *Escherichia coli* ATP synthase.";  
Nucleic Acids Res. 9:5387-5296 (1981).  
[3]  
SEQUENCE FROM N.A.  
SPECIES=E.coli;  
MEDLINE=82134798; PubMed=6277310;  
Kanazawa H., Kayano T., Mabuchi K., Futai M.;  
"Nucleotide sequence of the genes coding for alpha, beta and gamma  
subunits of the proton-translocating ATPase of *Escherichia coli*.";  
Biochem. Biophys. Res. Commun. 103:604-612 (1981).  
[4]  
SEQUENCE FROM N.A.  
SPECIES=E.coli;  
MEDLINE=83176724; PubMed=6301339;  
Kanazawa H., Futai M.;  
"Structure and function of H<sup>+</sup>-ATPase: what we have learned from  
*Escherichia coli* H<sup>+</sup>-ATPase".  
Ann. N. Y. Acad. Sci. 402:45-64 (1982).  
[5]  
SEQUENCE FROM N.A.  
SPECIES=E.coli; STRAIN=K12 / MG1655;  
MEDLINE=93315143; PubMed=7686882;  
Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;  
"DNA sequence and analysis of 136 kilobases of the *Escherichia coli*  
genome: organizational symmetry around the origin of replication";  
Genomics 16:551-561 (1993).  
[6]  
SEQUENCE FROM N.A.  
SPECIES=E.coli; STRAIN=O6:HL / CFT073 / ATCC 700928 / UPEC;  
MEDLINE=22388234; PubMed=14271157; DOI=10.1073/pnas.242529799;  
Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
"Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic *Escherichia coli*";  
Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
[7]  
SEQUENCE FROM N.A.  
SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;  
MEDLINE=21074935; PubMed=11206553; DOI=10.1038/35054089;  
Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
Fosf J., Hackett J., Plunkett G. III, Blattner F.R.,

SPCIES-E-2017; STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;  
MEDLINE=201704935; PubMed=11206551; DOI=10.1038/35054089;  
Perna N.T., Plunkett G. III, Burdland V., Mau B., Glasner J.D.,  
Rose D.J., Mayhew G.F., Evans F.S., Gregor J., Kirkpatrick H.A.,  
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
RL Nature 409:529-533(2001).  
RL [8]  
RP SEQUENCE FROM N.A.  
RC SPECIES=*E.coli*; STRAIN=O157:H7 / RIMD 0509952 / EHEC;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasaki A.C., Ogasawara C., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shihagawa H.;  
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
RL [9]  
RN SEQUENCE OF 261-287 FROM N.A.  
RP SPECIES=*E.coli*;  
RC SPECIES=*E.coli*;  
RX MEDLINE=90202983; PubMed=2138624;  
RA Iwamoto A., Miki J., Maeda M., Futai M.;  
RT "H(+)-ATPase gamma subunit of *Escherichia coli*. Role of the conserved  
RT carboxyl-terminal region.";  
RL J. Biol. Chem. 265:5043-5048(1990).  
RL [10]  
RN SEQUENCE OF 2-9; 72-81; 203-208 AND 214-220.  
RP SPECIES=*E.coli*;  
RC SPECIES=*E.coli*;  
RX PubMed=7508444;  
RA Tang C., Wilkens S., Capaldi R.A.;  
RT "Structure of the gamma subunit of *Escherichia coli* F1 ATPase probed  
RT in trypsin digestion and biotin-avidin binding studies.";  
RL J. Biol. Chem. 269:4467-4472(1994).  
RL [11]  
RN SEQUENCE FROM N.A.  
RP SPECIES=*S.flexneri*; STRAIN=301 / Serotype 2a;  
RC MEDLINE=2227406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
RT through comparison with genomes of *Escherichia coli* K12 and O157.";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
RL [12]  
RN SEQUENCE FROM N.A.  
RP SPECIES=*S.flexneri*; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RC MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of *Shigella*  
RT *flexneri* serotype 2a strain 2457T.";  
RL Infect. Immun. 71:2775-2786(2003).  
RL [13]  
RN X-RAY CRYSTALLOGRAPHY (4.4 ANGSTROMS).  
RP SPECIES=*E.coli*;  
RC MEDLINE=20040613; PubMed=10570135;  
RA Hausarth A.C., Gruber G., Matthews B.W., Capaldi R.A.;  
RT "Structural features of the gamma subunit of the *Escherichia coli* F(1)  
RT ATPase revealed by a 4.4-A resolution map obtained by X-ray  
RT crystallography.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:13697-13702(1999).  
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton  
CC gradient across the membrane. The gamma chain is believed to be  
CC important in regulating ATPase activity and the flow of protons  
CC through the CF(0) complex.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +  
CC H(+) (Out).  
CC -!- SUBUNIT: P-type ATPases have 2 components, CF(1) - the catalytic  
CC core and CF(0) - the membrane proton channel. CF(1) has five  
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)  
CC

CC has three main subunits: a, b and c.  
CC -!- SIMILARITY: Belongs to the ATPase gamma chain family.  
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts  
CC and various other errors.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X01631; CAA25781.1; -  
DR EMBL; J01594; AAA24736.1; ALT\_FRAME.  
DR EMBL; V00267; CAA23526.1; -  
DR EMBL; V00312; CAA23597.1; ALT\_FRAME.  
DR EMBL; M25464; AAA83874.1; -  
DR EMBL; L10328; AAA62085.1; -  
DR EMBL; AE000450; AAC76756.1; -  
DR EMBL; AE016769; AAH83091.1; -  
DR EMBL; AE005605; AAG58936.1; -  
DR EMBL; AP002566; BAB38098.1; -  
DR EMBL; M34095; AAA24742.1; -  
DR EMBL; AE015388; AAA45553.1; -  
DR EMBL; AE016991; AAP18944.1; -  
DR PIR; A01038; PWECC.  
DR PIR; C91213; C91213.  
DR PIR; D86059; D86059.  
DR PDB; 1D8S; X-ray; G=-.  
DR PDB; 1FS0; X-ray; G=19-248.  
DR EcoGene; EB0102; -  
DR InterPro; IPR000131; ATPase\_gamma.  
DR Pfam; PF00231; ATP-synt; 1.  
DR PRINTS; PRO0126; ATPASEGAMMA.  
DR TIGRFAMS; TIGR01146; ATPsyn\_Flgamma; 1.  
DR PROSITE; PS00153; ATPASE\_GAMMA; 1.  
DR 3D-structure; ATP synthetase; CF(1); Complete proteome;  
DR Direct protein sequencing; Hydrogen ion transport; Hydrolase.  
FT HELIX 20 57  
FT HELIX 66 68  
FT STRAND 75 81  
FT TURN 89 90  
FT TURN 91 108  
FT TURN 109 110  
FT STRAND 112 118  
FT HELIX 120 129  
FT STRAND 133 137  
FT TURN 141 142  
FT HELIX 147 161  
FT TURN 162 163  
FT STRAND 168 177

Query Match 69.3%; Score 991; DB 1; Length 287;  
Best Local Similarity 66.1%; Pred. No. 4.3e-57;  
Matches 191; Conservative 47; Mismatches 49; Indels 2; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASKTQTERMSRRPYSETIRNWIHVSXKAT 60  
Db 1 MAGAKEIRTKIASVQNTQKITKAMEMVAASKTQTERMSRRPYSETIRNWIHVSXKAT 60  
QY 61 IGYKHPLVDREVKKYGMIVYSTDRGLCGGLNVLNFKTVLNEKWEKEDVSVQLSLIGS 120  
Db 61 LEYKHPYLEDVRKVRGYLVVSTDRGLCGGLNVLNFKTVLNEKWEKEDVSVQLSLIGS 120  
QY 121 KSNIFPQSLGKILTDGSDIGTPSVEQLIGSVNGMIDAYKKGEVDVYVYVYKFNFTMS 180  
Db 121 KGVSFNFVSGNVVAVQVTGMDGNPFLSELIGPVMQLQAYDEGRDLKLYIVNKNFTMS 180  
QY 181 QKPVLEKLIPLPELDNDELGERKQVNDVIYEPDAKVLNLDNLLVRYLESQVTAAVENLAS 240  
Db 181 QVPTISQLPLPASDDDL--KHKSWDLYBEDPKALLDTLLRRYVESQVYQGVVENLAS 238

Qy 241 EQAARMVAMKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAI 289  
 Db 239 EQAARMVAMKAATDNGGSLIKELQLVYNKARQASITOEITEIVSGAAV 287

## RESULT 8

Q9RFL4	PRELIMINARY;	PRT; 287 AA.
ID	Q9RFL4	
AC	Q9RFL4;	
DT	01-MAY-2000	(TREMBLrel. 13, Created)
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)
DE	ATP synthase subunit gamma.	
DE	Name-atpG;	
GN	Salmonella typhimurium.	
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
OC	Enterobacteriaceae; Salmonella.	
OX	NCBI TaxID=602;	

RP SEQUENCE FROM N.A.

RC	STRAIN-TA98;
RRA	Kim H.-K., Heo N.-J., Ghim S.-Y., Song B.-H.;
RRL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DRR	EMBL; AF198265; AAF19361.1; .
DR	HSSP; P00837; 1FSO.
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0016469; F:proton-transporting two-sector ATPase complex; IEA.
DR	GO; GO:0046933; F:hydrogen-transporting ATP synthase activity. . ; IEA.
DR	GO; GO:0046961; F:hydrogen-transporting ATPase activity. . ; IEA.
DR	GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.
DR	InterPro; IPR00131; ATPase_gamma.
DR	Pfam; PF00231; ATP-synt; 1.
DR	PRINTS; PR00231; ATPSEGAMMA.
DR	TIGRFAMS; TIGR01146; ATPSyn_F1gamma; 1.
DR	PROSITE; PS00153; ATPASE_GAMMA; 1.
NO	SEQUENCE 287 AA; 31578 MW; AOB3E8D41B3B80F6 CRC64;

Query Match 69.3%; Score 991; DB 2; Length 287;  
 Best Local Similarity 56.1%; Pred. No. 4,3e-57;  
 Matches 191; Conservative 47; Mismatches 49; Indels 2; Gaps 1;

1	MAGAK	I	R	T	K	I	A	S	V	K	S	T	Q	K	I	T	A	M	E	N	V	A	A	S	K	U	R	K	T	O	E	R	M	S	S	R	P	P	S	E	T	I	R	N	V	I	S	H	V	S	K	A	T	60		
1	MAGAK	I	R	S	K	I	A	S	V	Q	N	T	Q	K	I	T	A	M	E	N	V	A	A	S	K	U	R	K	S	Q	D	R	M	A	S	R	P	P	E	T	I	R	N	V	I	S	H	V	S	K	A	T	60			
61	I	G	K	H	E	P	L	V	D	R	E	V	K	K	G	M	I	V	S	T	D	G	L	C	G	L	N	V	A	L	F	K	T	V	N	E	M	K	E	K	E	K	O	V	S	V	Q	L	S	I	G	S	120			
61	L	E	V	K	H	E	P	L	E	D	R	V	K	V	G	V	I	V	S	T	D	G	L	C	G	L	N	I	N	F	K	L	A	B	E	M	K	T	W	D	K	G	V	Q	C	D	L	A	M	I	G	S	120			
121	K	S	I	N	F	Q	S	L	I	K	I	L	T	Q	D	S	G	I	G	T	P	S	E	O	L	I	G	S	V	N	S	M	I	D	A	Y	K	K	G	E	V	D	V	V	V	I	V	N	K	F	I	N	T	M	S	180
121	K	G	V	S	F	D	S	V	G	N	V	A	Q	V	T	G	M	D	N	P	S	L	S	E	L	I	G	P	V	K	V	M	L	Q	A	Y	D	E	G	R	L	D	K	L	I	V	S	N	K	F	I	N	T	M	S	180
181	Q	K	P	V	L	E	K	I	P	E	L	N	D	E	L	G	E	R	K	O	W	N	D	I	Y	E	P	D	A	K	V	L	N	I	L	V	I	E	S	O	T															

## RESULT 9

Q8Z9S5 PRELIMINARY; PRT; 287 AA.  
 Q8Z9S6 Q74PA0; Q7CFW7;  
 Q8Z9S7 01-MAR-2002 (TREMBLrel. 20, Created)  
 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 01-OCT-2004 (TREMBLrel. 28, Last annotation update)  
 ATP synthase gamma subunit protein (EC 3.6.1.34) (Membrane-bound ATP  
 synthase, F1 sector, gamma-subunit).  
 Name:atpG; Synonyms:papC, uncG;

GN OS OC OC OX RN RP RC RX RA RA RA RA RA RA RA RT RL

SEQUENCE FROM N.A.  
STRAIN=KIMS / Biovar Mediaevalis;  
MEDLINE=22137863; PubMed=12142430;  
Deng W., Burland V., Plunkett G. III, Bout  
Perna N.T., Rose D.J., Mauck B., Zhou S., Sc  
Fetherston J.D., Lindler L.E., Brubaker R.  
RA Straley S.C., McDonough K.A., Nilles M.L.,  
RA Perry R.D.;  
RT "genome sequence of *Yersinia pestis* KIM." ;  
RN J. Bacteriol. 184:4601-4611(2002).  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=91001 / Biovar Mediaevalis;  
RC Song Y., Tong Z., Wang L., Han Y., Zhang J  
RA Han Y., Wang X., Zhai Y., Chen F., Qin H.,  
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang  
RA Yang R.;  
RA Submitted (APR-2003) to the EMBL/GenBank/D  
DB EMBL: AY141156

DR EMBL; AJ414160; CAC93571.1; -;  
DR EMBL; AE014015; AAM87678.1; -;  
DR EMBL; AE017142; AAS64168.1; -;  
DR EMBL; AF0500; AF0500.1; -;

PIR: AF0500; AF0500.  
HSP; P00837; IFS0.  
GO: GO:0016020; C:membrane; IEA.  
GO: GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.  
GO: GO:0046933; F:hydrogen-transporting ATP synthase activity; IEA.  
GO: GO:0046961; F:hydrogen-transporting ATPase activity, rota.; IEA.  
GO: GO:0016787; F:hydrolase activity; IEA.  
GO: GO:0015986; P:ATP synthetis coupled proton transport; IEA.  
InterPro: IPR000131; ATPase\_gamma.  
InterPro: IPR000005; HTHarac.  
Pfam: PF00231; ATP-synt.; 1.  
PRINTS; P000126; ATPSEGAMMA.  
TIGRFAMS; TIGR01146; ATPsyn\_Figamma; 1.  
PROSITE; PS00153; ATPASE\_GAMMA; 1.  
PROSITE; PS00041; HTH\_AEAC\_FAMILY\_1; UNKNOWN\_1.  
Complete proteome; Hydrolase.  
SEQUENCE 287 AA; 31577 MW; 2F124E8B7CE031CA CRC64;

Query Match	69.2%;	Score 989;	DB 2;	Length 287;
Best Local Similarity	67.1%;	Pred. No. 5.8e-57,		

[illegible]

RESULT 11	PRELIMINARY;	PRT;	287 AA.
D	Q8ZKW8		
D	Q8ZKW8;		
T	01-MAR-2002 (TREMBLrel. 20, Created)		
T	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
T	01-JUN-2003 (TREMBLrel. 24, Last annotation update)		
E	Membrane-bound ATP synthase, F1 sector, gamma-subunit.		
N	Name=atpbg; OrderedLocusNames=STM3866;		
N	Salmonella typhimurium.		
S	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
C	Enterobacteriaceae; Salmonella.		
C	NCBI_TaxID=602;		
N	[1]		

```

RESULT 12
Q8EBB9 Q8EBB9 PRELIMINARY; PRT; 286 AA.
ID ID
Q8EBB9 Q8EBB9;
AC AC
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ATP synthase F1, gamma subunit.
GN Name=atpG; OrderedLocuName=SO4748;
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI TaxID=70863;
RN [1]_
RPN RN
RRP SEQUENCE FROM N.A.
RRC STRAIN=MR-1;
RRE MEDLINE=22297686; PubMed=12369813; DOI=10.1038/nbt749;
RRA Haddelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RRA Read T.D., Eisen J.A., Seshadri R., Ward N.L., Meehe B.A.,
RRA Clayton R.A., Meyer T., Teapin A., Scott J., Beanan M.J.,
RRA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RRA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,

```

RA	White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Imbrahim M.,
RA	Lee K., Berry K.J., Lee C., Mueller J., Khouri H.M., Gill J.,
RA	Utterback T.R., McDonald L.A., Feldblum T.V., Smith H.O.,
RA	Venter J.C., Nealson K.H., Fraser C.M.,
RT	"Genome sequence of the dissimilatory metal ion-reducing bacterium
RT	Shewanella oneidensis",
RL	Nat. Biotechnol. 20:1118-1123 (2002).
DR	EMBL; AE015907; AAN57707.1; -
DR	HSSP; P00837; IFSO.
DR	TIGR; SO4748; -
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
DR	GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; -; IEA.
DR	GO; GO:0046361; F:hydrogen-transporting ATPase activity, rota. -; IEA.
DR	GO; GO:0015386; P:ATP synthesis coupled proton transport; IEA.
DR	InterPro; IPR00131; ATPase_gamma.
DR	Fam; PF00231; ATP-synt; 1.
DR	PRINTS; PR00126; ATPASEGAMMA.
DR	TIGRFAMS; TIGR01146; ATPsyn_F1gamma; 1.
DR	PROSITE; PS00153; ATPASE_GAMMA; 1.
KW	Complete proteome.
SQ	SEQUENCE 286 AA; 31515 MW; B3549DB862059053 CRC64;
Query Match 68.8%; Score 982.5; DB 2; Length 286;	
Best Local Similarity 64.4%; Pred. No. 1.5e-56;	
Matches 186; Conservative 55; Mismatches 45; Indels 3; Gaps 1;	
QY	1 MAGAKEIRTKIASVKSTOKITKAMENWAASKVKTKQERMSSRPYSYETIRNVISHSKAT 60
DB	1 MAGAKEIKTKLASVKNVTKIYSAMENWAASKVRRAQERMAASRPYASMRKVIGHVAQGS 60
QY	61 IQYKHPFLVDREVKKVGMIVYSTDRGLCGGLNVNLFTVLNEMKEWKEKDVSVQLSLIGS 120
DB	61 LEYKHPYLEVREAKTEGVYIVVATDRGLCGGLNVNLFKKVADVKSKEQGAEEFCPIGA 120
QY	121 KSINFFQSILGKILTDGSGIGDTSFVOLIGSVNSMIDAYKKGEVDVYLVYVNFINTMS 180
DB	121 RSVQFFKSGCGQAAGSLGDAPKINDLITGVQWLEAYNEGLDRLYVVFNFVNTWT 180
QY	181 QKPVLEKLIPLPELDNDLDELGERKQWDYIYEPDAKVLIDNLLVYVLESQYQAENVLAS 240
DB	181 QTPVIEQLLPPLKSEDDVAHR---WDYIYEPDKALLDILLVYVYESQYQGWENIAS 237
QY	241 EQAARMVAMKAATNDAGNLINELQLVYNKARQASITNELNEIVAGAAAI 289
DB	238 EQAARMVAMKAATNDAGTLDDIDLQLVYNKARQAATQELSEIVSGASAV 286
RESULT 13	
Q7MGH9	PRELIMINARY; PRT; 288 AA.
ID	Q7MGH9
AC	Q7MGH9;
DT	01-MAR-2004 (TrEMBLrel. 26, Created)
DT	01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE	FOR1-type ATP synthase, gamma subunit.
OS	Names=VJ3252;
GN	Vibrio vulnificus (strain VJ016).
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC	Vibrionaceae; Vibrio.
OX	NCBI_TaxID=196600;
RX	[1]
RP	SEQUENCE FROM N.A.
RA	PubMed=14656965;
RA	Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA	Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA	Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT	"Comparative genome analysis of Vibrio vulnificus, a marine
RT	pathogen,"
RL	Genome Res. 13:2577-2587 (2003).
DR	EMBL; AF005343; BAC96016.1; -
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.

```

QY 1 MAGAKEIRTKIASVKTOKITKAMEMVAASKMKTKQERMSSSRPYSETIRNVISHVSKAT 60
Db 1 MAGAKEIRSKIGSVKSTOKITKAMEMVAASKMKTKQERMSSSRPYSETIRNVISHVSKAT 60
QY 61 IGYKHPFLVDREVKKVGMIWSTDRGLCGGLNVLNFKTVLNMKEWKEKDVSVQLSLIGS 120
Db 61 LEYRHPFLVDREVKKVGMIWSTDRGLCGGLNVLNFKTVLNMKEWKEKDVSVQLSLIGS 120
QY 121 KSNFPOSLGKILQDSGIGTTPSVQELIGSVNSMIDAYKKGVDVYVLYVYNKFINTMS 180
Db 121 KATAPFKHGAAGAAQVSGGLGSPSLDLIGSVVLEKYDEGLDRLYLVLNKFVNTMV 180
QY 181 QKPVLEKLIPLPELDNDELGERKQVWDYIYEPDAKVLNLLVLYLESQVYQAAVENLAS 240
Db 181 QOPTIDQLPLPKSDSKDM-QREHSDWYIYEPDAKVLNLLVLYLESQVYQAAVENLAS 240
QY 241 EQAARMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289
Db 240 EQAARMVAMKAATDNATNLDDLELVYNKARQASITNELNEIVAGAAAV 288

RESULT 15
Q822Q5
ID Q822Q5 PRELIMINARY; PRT; 287 AA.
AC Q822Q5; Q7C618;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE ATP synthase gamma subunit.
GN Name=atpG; OrderedLocusNames=STV3912, t3653;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21534947; PubMed=11677608; DOI=10.1038/35101607;
RX MEDLINE=21534947;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogan A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative Genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
DR EMBL; AL627280; CAD03129.1; -
DR EMBL; AE016846; AAO1150.1; -
DR HSSP; P00837; IFS0.
DR GO; GO:0046020; C:membrane; IEA.
DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; IEA.
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity; IEA.
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.
DR InterPro; IPR000131; ATPase_gamma.
DR Pfam; PF00231; ATP-synt; 1.
DR PRINTS; PR00126; ATPASEGAMMA.
DR TIGRFAMs; TIGR01146; ATPsyn_Figamma; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; 1.
KW Complete proteome.

```

```

SQ SEQUENCE 287 AA; 31475 MW; 2FB82E039979589 CRC64;
Query Match 68.4%; Score 978; DB 2; Length 287;
Best Local Similarity 64.7%; Pred. No. 3.1e-56;
Matches 187; Conservative 51; Mismatches 49; Indels 2; Gaps 1;
QY 1 MAGAKEIRTKIASVKTOKITKAMEMVAASKMKTKQERMSSSRPYSETIRNVISHVSKAT 60
Db 1 MAGAKEIRSKIGSVKSTOKITKAMEMVAASKMKTKQERMSSSRPYSETIRNVISHVSKAT 60
QY 61 IGYKHPFLVDREVKKVGMIWSTDRGLCGGLNVLNFKTVLNMKEWKEKDVSVQLSLIGS 120
Db 61 LEYRHPFLVDREVKKVGMIWSTDRGLCGGLNVLNFKTVLNMKEWKEKDVSVQLSLIGS 120
QY 121 KSNFPOSLGKILQDSGIGTTPSVQELIGSVNSMIDAYKKGVDVYVLYVYNKFINTMS 180
Db 121 KATAPFKHGAAGAAQVSGGLGSPSLDLIGSVVLEKYDEGLDRLYLVLNKFVNTMV 180
QY 181 QKPVLEKLIPLPELDNDELGERKQVWDYIYEPDAKVLNLLVLYLESQVYQAAVENLAS 240
Db 181 QOPTIDQLPLPKSDSKDM-QREHSDWYIYEPDAKVLNLLVLYLESQVYQAAVENLAS 240
QY 241 EQAARMVAMKAATDNAGNLINELQVYNKARQASITNELNEIVAGAAAI 289
Db 240 EQAARMVAMKAATDNATNLDDLELVYNKARQASITNELNEIVAGAAAV 288

Search completed: November 9, 2004, 09:56:59
Job time : 197 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2004, 09:48:14 ; Search time 39 Seconds  
(without alignments)  
491.433 Million cell updates/sec

Title: US-09-545-199F-4  
Perfect score: 1429  
Sequence: 1 MAGAKEIRTKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/6C\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1429	100.0	289	4	US-09-809-665A-4
2	1132.5	79.3	288	4	US-09-809-665A-133
3	1077.5	75.4	288	4	US-09-809-665A-167
4	1004	70.3	291	4	US-09-543-681A-7449
5	978	68.4	297	4	US-09-489-039A-10616
6	926.5	64.8	291	4	US-09-252-991A-19080
7	894	62.6	292	4	US-09-328-352-4894
8	885	61.9	309	4	US-09-540-236-2848
9	530.5	37.1	304	4	US-09-134-000C-5902
10	524.5	36.7	289	4	US-09-254-504-4
11	510.5	35.7	300	4	US-09-107-532A-7284
12	504.5	35.3	289	4	US-09-254-504-8
13	501.5	35.1	292	4	US-09-583-110-3385
14	493	34.5	295	3	US-09-134-001C-4242
15	464	32.5	272	4	US-09-710-279-1812
16	460.5	32.2	293	4	US-09-254-504-12
17	440	30.8	308	4	US-08-311-731A-148
18	231.5	16.2	281	4	US-09-861-451A-22
19	179	12.5	183	4	US-09-248-796A-17416
20	143.5	10.0	139	4	US-09-513-999C-4941
21	136.5	9.6	136	4	US-09-270-767-32542
22	136.5	9.6	136	4	US-09-270-767-47759
23	108.5	7.6	383	4	US-09-328-352-7081
24	102.5	7.2	10182	3	US-09-134-001C-3159
25	102	7.1	386	2	US-08-455-968E-7
26	100	7.0	721	4	US-09-248-796A-15107
27	97	6.8	289	4	US-09-107-532A-4402

28	95	6.6	430	4	US-09-328-352-5624	Sequence 5624, Ap
29	94	6.6	487	4	US-09-107-532A-5731	Sequence 5731, Ap
30	93.5	6.5	718	3	US-08-973-005A-12	Sequence 12, Appl
31	92	6.4	315	3	US-08-793-035-9	Sequence 9, Appl
32	92	6.4	315	3	US-08-793-035-10	Sequence 10, Appl
33	92	6.4	775	1	US-07-603-133B-16	Sequence 16, Appl
34	92	6.4	786	4	US-09-248-796A-15288	Sequence 15288, A
35	91.5	6.4	635	4	US-09-538-092-1	Sequence 1, Appl
36	91.5	6.4	2349	4	US-08-480-604A-6	Sequence 914, Ap
37	91.5	6.4	2710	1	US-08-480-604A-6	Sequence 6, Appl
38	91.5	6.4	2710	2	US-08-405-496A-6	Sequence 6, Appl
39	91.5	6.4	2710	3	US-08-915-136-6	Sequence 6, Appl
40	91.5	6.4	2710	3	US-08-957-310-6	Sequence 6, Appl
41	91.5	6.4	2710	4	US-10-011-366-6	Sequence 6, Appl
42	91.5	6.4	2710	4	US-09-084-517-6	Sequence 6, Appl
43	91.5	6.4	3878	4	US-09-914-259-11	Sequence 11, Appl
44	91	6.4	444	4	US-09-248-796A-19173	Sequence 19173, A
45	91	6.4	1010	3	US-09-134-001C-5178	Sequence 5178, Ap

## ALIGNMENTS

RESULT 1  
US-09-809-665A-4  
; Sequence 4, Application US/09809665A  
; Patent No. 6790950  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809,665A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 09/545,199  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Pasteurella multocida  
US-09-809-665A-4

Query Match	100.0%;	Score 1429;	DB 4;	Length 289;
Best Local Similarity	100.0%;	Pred. No. 2.4e-136;		
Matches 289;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAGAKEIRTKIASVKSTOKITKAMEWVAASVKRKTQREMSRRPYSETIRNVISHVSKAT	60	
Db	1	MAGAKEIRTKIASVKSTOKITKAMEWVAASVKRKTQREMSRRPYSETIRNVISHVSKAT	60	
QY	61	IGYKHPFLVDREVKVGMIVVSTDRGLCGGLNNLFTVLNEMKEWKEKDVSVQLSLIGS	120	
Db	61	IGYKHPFLVDREVKVGMIVVSTDRGLCGGLNNLFTVLNEMKEWKEKDVSVQLSLIGS	120	
QY	121	KSINFFQSLGKILITQDSGIGDTPSVQQLIGSVNSMIDAYKKEVDVYLVYVKPNTMS	180	
Db	121	KSINFFQSLGKILITQDSGIGDTPSVQQLIGSVNSMIDAYKKEVDVYLVYVKPNTMS	180	
QY	181	QKPVLEKLIPLPDLNDDELGERKQVWDYIYEPDAKVLNDLLVRYLESQVYQAAVENLAS	240	
Db	181	QKPVLEKLIPLPDLNDDELGERKQVWDYIYEPDAKVLNDLLVRYLESQVYQAAVENLAS	240	
QY	241	EQAAWVAMKAATDNAGNLINELQLVYVKARQASITNELNEIVAGAAAI	289	
Db	241	EQAAWVAMKAATDNAGNLINELQLVYVKARQASITNELNEIVAGAAAI	289	

RESULT 2



US-09-809-665A-133  
; Sequence 133, Application US/09809665A  
; Patent No. 6790950  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809,665A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 09/545,199  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 133  
; LENGTH: 288  
; TYPE: PRT  
; ORGANISM: Actinobacillus pleuropneumoniae  
US-09-809-665A-133

Query Match 79.3%; Score 1132.5; DB 4; Length 288;  
Best Local Similarity 74.4%; Pred. No. 2.4e-106;  
Matches 215; Conservative 45; Mismatches 28; Indels 1; Gaps 1;

Qy 1 MAGAKEIRTKIASVSKSTQKITKAMENVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Db 1 MAGAKEIRTKIASVSKSTQKITKAMENVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Qy 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKQVSVOLSLIGS 120  
Db 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKQVSVOLSLIGS 120  
Qy 121 KSNFPQSLGKILTDQSGIGDTPSEVOLIGSVNSMIDAYKKGEVDVYLVYVYVNFINTMS 180  
Db 121 KSNFPQSLGKILTDQSGIGDTPSEVOLIGSVNSMIDAYKKGEVDVYLVYVYVNFINTMS 180  
Qy 181 QKPVLEKLIPLPELNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240  
Db 181 QKPVLEKLIPLPELNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240  
Qy 241 EQAARMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAAI 289  
Db 241 EQAARMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAAI 289

RESULT 3  
US-09-809-665A-167  
; Sequence 167, Application US/09809665A  
; Patent No. 6790950  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809,665A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 09/545,199  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 167  
; LENGTH: 288  
; TYPE: PRT  
; ORGANISM: Pasteurella (Mannheimia) haemolytica  
US-09-809-665A-167

Query Match 75.4%; Score 1077.5; DB 4; Length 288;  
Best Local Similarity 70.3%; Pred. No. 2.5e-93;  
Matches 194; Conservative 49; Mismatches 44; Indels 2; Gaps 1;

Best Local Similarity 72.0%; Pred. No. 8.9e-101;  
Matches 208; Conservative 46; Mismatches 34; Indels 1; Gaps 1;  
Qy 1 MAGAKEIRTKIASVSKSTQKITKAMENVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Db 1 MAGAKEIRTKIASVSKSTQKITKAMENVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Qy 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKQVSVOLSLIGS 120  
Db 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKQVSVOLSLIGS 120  
Qy 121 KSNFPQSLGKILTDQSGIGDTPSEVOLIGSVNSMIDAYKKGEVDVYLVYVYVNFINTMS 180  
Db 121 KSNFPQSLGKILTDQSGIGDTPSEVOLIGSVNSMIDAYKKGEVDVYLVYVYVNFINTMS 180  
Qy 181 QKPVLEKLIPLPELNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240  
Db 181 QKPVLEKLIPLPELNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240  
Qy 241 EQAARMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAAI 289  
Db 241 EQAARMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAAI 289

RESULT 4  
US-09-543-681A-7449  
; Sequence 7449, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7449  
; LENGTH: 291  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-7449

Query Match 70.3%; Score 1004; DB 4; Length 291;  
Best Local Similarity 67.1%; Pred. No. 2.5e-93;  
Matches 194; Conservative 49; Mismatches 44; Indels 2; Gaps 1;

Qy 1 MAGAKEIRTKIASVSKSTQKITKAMENVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Db 5 MAGAKEIRTKIASVSKSTQKITKAMENVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Qy 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKQVSVOLSLIGS 120  
Db 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKQVSVOLSLIGS 120  
Qy 121 KSNFPQSLGKILTDQSGIGDTPSEVOLIGSVNSMIDAYKKGEVDVYLVYVYVNFINTMS 180  
Db 121 KSNFPQSLGKILTDQSGIGDTPSEVOLIGSVNSMIDAYKKGEVDVYLVYVYVNFINTMS 180  
Qy 181 QKPVLEKLIPLPELNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240  
Db 181 QKPVLEKLIPLPELNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240  
Qy 241 EQAARMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAAI 289  
Db 241 EQAARMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAAI 289

RESULT 5  
US-09-489-039A-10616  
; Sequence 10616, Application US/09489039A  
; Patent No. 6610836

```

; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10616
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-10616

```

Query Match	68.4%;	Score 978;	DB 4;	Length 297;
Best Local Similarity	65.1%;	Pred. No. 1.1e-90;		
Matches 188;	Conservative 50;	Mismatches 49;	Indels 2;	Gaps 1;
QY	1	MAGAKEIRTKIASVKSTOKITKAMEMVASKKVEKTOERSSSRPSPYSEIRNVISHVSKAT	60	
Db	11	MAGAKEIRSKIASVQNTQKITKAMEMVASKKVEKTOERSSSRPSPYSEIRNVISHVSKAT	70	
QY	61	IGYKHPILVDREKVKGMIVVSDTRGLCGGLNVNLFKTVLNMKEWKEDKYSVSLSTIGS	120	
Db	71	LEYKHPILVEERDVKRGVYLVSDTRGLCGGLNVNLFKTLAEMKAWSDKGVQCDLAMI	130	
QY	121	KSINFFOSLGIKILTDOSGIGDTPSYEQVLIIGSVNSMIDAYKKGEVDVYVLYVYKFNITMS	180	
Db	131	KGVSFFNSVSGNVVAQVTGMGNPNLSLELIGPVKWLQAYDEGRDLKLYVYVYKFNITMS	190	
QY	181	QKPVLEKILPELNDDELGERKQVWDVYIYEDPAKVILNDLNLVRYLSQVTOAAVEMLAS	240	
Db	191	QVPTITQLLELPASEADL--RRKSWDYIEDPDKALLDTLLRYVTSQVYQGVVENLAS	248	
QY	241	EQARVMVAKAATDNAGTNINELQVLVYKCAQASITNEINEIVAGAAAI	289	
Db	249	EQARVMVAKAATDNAGSILKELQVLVYKCAQASITQELTEIVSGAAAV	27	

```

RESULT 6
US-09-252-991A-19080
; Sequence 19080, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19080
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19080

```

	Query Match	64.8%;	Score 926.5;	DB 4;	Length 291;	
	Best Local Similarity	62.3%;	Pred. No. 1.7e-85;			
	Matches 180;	Conservative 50;	Mismatches 56;	Indels 3;	Gaps 17;	
OY	1	MAGAEIRTKIASVKSTQKTIKAMEVAAASKTGERMSSSRPYSSETIRNVISHVSTAT	60			
		:::::::::::::::::::::	:	:	:	:
db	6	MAGAEIRSKIASIKSTQKITNAMEKVAVSQRKAQVRMAGRPRVAERIQVIGHLANAN	65			
		:::::::::::::::::::::	:	:	:	:
OY	61	IGYKHPLVDREVKVGMIYVSITRGLCGGLNNVLPKTVLNMKEWKCKDVSVOQLSIGS	120			
		:::::::::::::::::::::	:	:	:	:

D <b>b</b>	66	PEYRHPFMYVEREKVGYIYWSSDRGLCGGLNINLFKSLVKDMSGYEQGAIEDLCWIGS	125
Q <b>y</b>	121	KSINFFQSLGIKILQDSIGDTSVEQOLIGSVNSMDAYKKEVDVYVLYNKFINMTMS	180
D <b>b</b>	126	KGASPRSPGGNVAAISHLGEBSINDLIGSVKMILDVLEGRIDFLFVSVKNFVNTWT	185
Q <b>y</b>	181	QKPVLEKLIPELDNDELGBERKQWMDYIPEYDAKVLDDMLLVYLESVQVYQAAVENLAS	240
D <b>b</b>	186	QKPTVEQLIPLVADDQDEL--KHHWDYIPEYDAKSLDGLLVRYVESQVYQAVVNNAC	242
Q <b>y</b>	241	EQGAARWAKAATONAGNLINELQIVYNKARQASITNMLNEIVAGAAAI	289
D <b>b</b>	243	EQGARVYANKRATDNAGELISDLGLIYNKARQAITQEI SEIVGGAAAV	291

```

RESULT 7
US-09-328-352-4894
; Sequence 4894, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4894
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4894

```

Query Match	62.6%	Score	894	DB 4	Length	292	
Best Local Similarity	59.7%	Pred.	No. 3.4e-82				
Matches	173	Conservative	56	Mismatches	59	Indels	2
2Y	1	MAGAKEIRTKIASVKSTOKITKAMENWVASKMKKTCQERMSRSPYSETIRNVISHVSKAT	60				
Db	4	MANLKEIRAKVASIKSTOKITPANQVAAKSNRRACERMAQGRPYADNRRRVIAHLVQAN	63				
2Y	61	IGYKHPLFDYREVKKVGMIVSTDRLGCGGLNVNLFKTVLNMKMKWKDKVSVLSLIGS	120				
Db	64	PEYKHRYWDRPVKRVGYIIVSSDRGLAGGINLFFKKVQHVHKAQOEQSIYQFALLIGQ	123				
2Y	121	KSINFPQSGLIKILTQDSIGIGTSPVEQLIGSVNSMIDAYKKGVGDVVYLVYNNKFINMS	160				
Db	124	KAVSFPKPYNGGKVLGATTQIGDAPSLQLTGSVQMLDAPDKGBELDRIVLVSGFVFNAMT	183				
2Y	181	QKPVLEKILPL-PELNDDELGERQVMDYIYEPDPAVLNDNLVRYLESQVYQAAVENLA	239				
Db	184	QKPKVEQLVPLAPAEGBDL-NRTYGMWYIYEPPEABELGLNVRIESVMYQGVNIENVA	242				
2Y	240	SECAARMVAMKAATDNAGLINELQLVYNKARQASITNELNEIVAGAAAI	289				
Db	243	CQSQARMVAMKAATDNAGLTKDLQLYNNKLROAAITQEIISIVGGAAAV	292				

RESULT 8  
 US-09-540-236-2848  
 ; Sequence 2848, Application US/09540236  
 ; Patent No. 6673910  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRI  
 ; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709-2005-001  
 ; CURRENT APPLICATION NUMBER: US/09/540,236  
 ; CURRENT FILING DATE: 2000-04-04  
 ; NUMBER OF SEQ ID NOS: 3840  
 ; SEQ ID NO 2848  
 ; LENGTH: 309  
 ; TYPE: PRT

ORGANISM: M. catarrhalis  
US-09-540-236-2848

Query Match  
Best Local Similarity 61.9%; Score 885; DB 4; Length 309;  
Matches 164; Conservative 63; Mismatches 62; Indels 0; Gaps 0;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASKORKTQERMSSRPYSRTIRNVISHVSKAT 60  
DB 20 NASLKEIRAKVTSIKTSQKITRAMQVMAASKORRAQERMELGRPYSGIRRVISHLVQAQ 79

QY 61 IGYPKPELVDRVKKVGMIVVSTDRGLCGGLVNLVFKTVLNEMKWKEDVSVQLSLIGS 120  
DB 80 SDYKHPYMINFVRKGVFVVISDRGLAGLNILFNKLLKTVKSQEQVEIEFPIGA 139

QY 121 KSNFFQSLGKILTDQSGIGDTPSVQQLIGSVNSMIDAYKKGEDVVVLVYNKFTINTMS 180  
DB 140 KGVGFKNFGGRVISAFTDYGDNPALEQINTPVQTMDDYINGKLDRIYLVYNQFINAMA 199

QY 181 QKPVLEKILPELDNDELGERKQWVDYIPEPDAKVLNLLVRYLESQVYQAAVENLAS 240  
DB 200 QKPVVEQIVPLAESEFDTETELQASHWDYIIEPDKTLDLSLLRYIESVYQVRENIA 259

QY 241 EQAARMVAMKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAAI 289  
DB 260 EQSARMVAMKAATDNAGNLKDLQLVYNKLFQAAITREISEIVGGAHV 308

RESULT 9  
US-09-134-000C-5902  
; Sequence 5902, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5902  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5902

Query Match  
Best Local Similarity 37.1%; Score 530.5; DB 4; Length 304;  
Matches 112; Conservative 68; Mismatches 106; Indels 15; Gaps 2;

QY 2 AGAKEIRTKIASVKSTQKITKAMEMVAASKORKTQERMSSRPYSRTIRNVISHVSKATI 61  
DB 5 ASLNEIKQRIASVTSKTSQKITKAMQVMAAKLTSEKASKSFQYSSKIRSVVTHLVAQ 64

QY 62 -----GVKHPFLVDREVKKVGMIVVSTDRGLCGGLVNLVFKTVLNEMKWK 108  
DB 65 SELRETEOSSISEGNYHVMKARPVKTKGYIVTSKGLVGYNSILKQTMSTQEDHD 124

QY 109 KDVSQLSLIGSKSINFFQSLGKILTDQSGIGDTPSVQQLIGSVNSMIDAYKKGEDVV 168  
DB 125 SNKEVALLAIGTGADFPKARGIDVSVYELRGLTDQPTFEVVRKIVTATTWQNEVFDEL 184

QY 169 YLVYNKFTINTMSQKPVLEKILPELDNDELGERKQWVDYIPEPDAKVLNLLVRYLES 228  
DB 185 YVCYNHHVNSUTSFRVEKMLPITDLPSEATSYEQ--EYLLEPSPEALDQLLPQYAES 242

QY 229 QVYQAAVENLASQQAARMVAMKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAA 288  
DB 243 LIYGLAIDAKTAHAAGTANKATDNAGNLISDLISYNARQASITQETIIVAGAAA 302

QY 289 I 289  
DB 303 L 303

RESULT 10  
US-09-254-504-4  
; Sequence 4, Application US/09254504  
; Patent No. 6511836  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: A method of improving the production of biomass  
; TITLE OF INVENTION: or a desired product from a cell  
; NUMBER OF SEQUENCES: 17  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA: US/09/254,504  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 963/96  
; FILING DATE: 06-SEP-1996  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 289 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-254-504-4

Query Match  
Best Local Similarity 36.7%; Score 524.5; DB 4; Length 289;  
Matches 107; Conservative 73; Mismatches 105; Indels 5; Gaps 3;

QY 2 AGAKEIRTKIASVKSTQKITKAMEMVAASKORKTQERMSSRPYSRTIRNVISHVSKAT 60  
DB 3 ASLNEIKQRIASVTSKTSQKITKAMQVMAAKLTSEKASKSFQYSSKIRSVVTHLVAQ 62

QY 61 IGYPKPELVDRVKKVGMIVVSTDRGLCGGLVNLVFKTVLNEMKWKEDVSVQLSLIGS 120  
DB 63 EPANQPMWIKREVKTGYLVTSDRGLVGSYNSNLKSVISNIRKHTNSEYITLALGG 122

QY 121 KSNFFQSLGKILTDQSGIGDTPSVQQLIGSVNSMIDAYKKGEDVVVLVYNKFTINTMS 180  
DB 123 TGADFPKARNVSVYVLRGLSDQPTFEVVRKIVTAEVEYQAEEDVLYVCYNHHVNSLV 182

QY 181 QKPVLEKILPELDNDELGERKQWVDYIPEPDAKVLNLLVRYLESQVYQAAVENLA 239  
DB 183 SEARMEKMLP---ISFDEKGEKASLVTFELEDRETILNQLLPQYAESMTYGSIVDAKT 239

QY 240 SEQAARMVAMKAATDNAGNLINELQLVYNKARQASITNELNEIVAGAAAI 289  
DB 240 AEHAAGTANKATDNAGNSVINDLTIQYNARQASITQETIIVAGASAL 289

RESULT 11  
US-09-107-532A-7284  
; Sequence 7284, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A. Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA

RESULT 12  
 US-09-254-504-8  
 Sequence 8, Application US/09254504  
 Patent No. 6511936  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: A method of  
 TITLE OF INVENTION: or a desired  
 NUMBER OF SEQUENCES: 17  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible

```

RESULT 13
US-09-583-110-3385
; Sequence 3385, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: Nucleic Acid and
; TITLE OF INVENTION: Pneumoniae for D
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US/9/107,4
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,1
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,5
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3385
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3385

```

```

Query Match      35.1%; Score 501.5; DB 4; Length 292;
Best Local Similarity 35.3%; Pred.No 1.8e-42;
Matches 101; Conservative 78; Mismatches 104; Indels 3; Gaps 2;

6 EIRTKIASVKSTQKITKAEWMAASKORIKTQRMSSRPYSYSETIRNVISHV--SKATIGY 63
7 DIKTKTASTKNTSQITINAVQWYSAAKLGEAEAFNFQVYACKVKLKITLHNGAGAS 66

```

Qy	64	KHPFLVDRVKVGMIVSVSTDRGCGGLNNVNFVTVLNEMKWKKBKDVSVQLSLIGSKSI	123
Db	67	TNPMLTISRVSVKGTGIVITSDRGLVGGYNSSILKAVNMLKEEYHPDGGPFGMIGGMA	126
Qy	124	NFFQSILGKILTKQDSGIGDTSPVQLIGSVNSMDAYKKGEVDVYVLVYNKFNNTMSQRP	183
Db	127	DFPKARGIQPLVELRGLADQPSFDQVRKILISKTVMQNELFDLVYVCYNHHVNTLTSQM	186
Qy	184	VLEKLTIPLEPNDBELGRKQWVDYIYBPDKAVLLDNLLVRYLESQVYQAAVENLASEQA	243
Db	187	RVEQMLPIVDLPDNEADSEYSLTFPELTSREEILEQLLPQFAESMIYGAIIDAKTAENA	245
Qy	244	ARMVAKAATDNAGNLINELQVLNFKARQASITNELNEIVAGAAAI	289
Db	246	AGMTAMQATDNNAKVIINDLTIQVNRARQAAITQITETIIVAGASAL	291

RESULT 14

```

US-09-134-001C-4242
; Sequence 4242, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4242
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4242

```

Query Match	34.5%	Score	493	DB 3	Length	295			
Best Local Similarity	33.2%	Pred. No.	1.4e-41						
Matches	96	Conservative	81	Mismatches	110	Indels	2	Gaps	1
QY	1	MAGAKERTKIASVKSTQKTKAMEMVAASQMKTKTQERMSSSRPSETIRNVISHVSKAT	60						
Db	8	MASLKEIDSRKTSKMKQITKAMNVSSSKLPRAEKTKSPRPTWKQDAITAVAGSN	67						
QY	61	IGYKHPFLVDREVTKVGMIVVSTDRLCGGLNVNLFKTVLNMKEWKEKDVSVQLSJLS	120						
Db	68	STSNHPMLKSDIIKRSGLYVITBCKGLAGAYSTNWLKSLVNDINSKHNDSEYSLIVLQ	127						
QY	121	KSINFFOSLGIKILTDQSGIGDTPSPVEQLIGSVNSMIDAYKKGEVDVVLVYNKFINMTS	180						
Db	128	QGVDFFKHGRGVEISSLVEVPDQSPFKSIQSIKHAIDLFPSENIDELTIYVSHVSVLE	187						
QY	181	QKPVLEKLIPLPFLNDDELGERQVWDYIYEPDAKVLNDLNLVRYLESQVYQAAVENLAS	240						
Db	188	NKPATKQVLPVLSQDSQ--GHQMSGSYEFEPDKSEILSVILPQVBSLIYGTITLDAKAS	245						
QY	241	EQAARWAMKAANTDAGNLNINELQLVYNKARQASITNINELNEIVAGAAAI	289						
Db	246	EHASRMTAMRNASNAPELIDDLSEYNKARQAAITOOITEIVGSSSL	294						

## RESULT 15

RESOL- 15  
US-09-710-279-1812  
; Sequence 1812, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMBERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EP  
; FILE REFERENCE: PU348005  
; CURRENT APPLICATION NUMBER: US/09/710

```

, CURRENT FILING DATE: 2000-11-09
, PRIOR APPLICATION NUMBER: 60/164,258
, PRIOR FILING DATE: 1999-11-09
, NUMBER OF SEQ ID NOS: 4472
, SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 1812
, LENGTH: 272
, TYPE: PRT
, ORGANISM: Artificial Sequence
, FEATURE:
, OTHER INFORMATION: Description of Artificial Sequence: synthetic
, OTHER INFORMATION: amino acid sequence
US-09-710-279-1812

```

Query Match	32.5%	Score 464;	DB 4;	Length 272;	
Best Local Similarity	32.7%;	79;	Midmatch 102;	Indels 2;	Gaps 1;
Matches	89;	Conservative	1;		
QY	18	QKITKAMEMVAASKORIKTORMSSSSPPYSETIRNVISHVSKATIGYKHPFLVDREVKKVG	77		
Db	2	QKITKAMWVSSSKLRAERNTKSPRYMEKWDAITAVAGSNSNSHPMLKSRDIKESG	61		
QY	78	MIVVSTDRGLCGGLNVLPFTVLNENKWEKXDVSQVLSIGSKSINFPQSLGKILQTQD	137		
Db	62	YLVITSDKGLAGVSTNVLSKLVNDINSKPNSSSEYSLVLGQQGVDFPKRGHYEIESL	121		
QY	138	SGIGTSPVEQIGSVNSMIDAYKKGEVDVVLVYKNKFINTMRSQPVLEKLIPLPELDND	197		
Db	122	VEVPDQSFKSIQSIAXHAIDLFEENIDELTIYSHVSVLENKPAKQVLPPLSQEDSG	181		
QY	198	ELGERKQWMDYIPEPDAKVLLDNLVYLVLESQVYQAAVENLASQAAARWAKAATONAG	257		
Db	182	Q--GHGQSSSYEFEPDKESLSVLFPQYVESUTYGTILDAKASEHASNRTAWRNASONAT	239		
QY	238	NLINELOLVYNKARQASITNELNEIVAGAAAI	289		
Db	240	ELIIDLLEYNARQAAITQIITEIVGGSSAL	271		

Search completed: November 9, 2004, 09:57:42  
Job time : 40 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2004, 09:50:44 ; Search time 140 Seconds  
(without alignments)  
723.159 Million cell updates/sec

Title: US-09-545-199F-4

Perfect score: 1429

Sequence: 1 MAGAKEIRTKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1429	100.0	289	11	US-09-809-665A-4
2	1132.5	79.3	288	11	US-09-809-665A-133
3	1077.5	75.4	288	11	US-09-809-665A-167
4	1007	70.5	287	14	US-10-369-493-21244
5	1002	70.1	287	14	US-10-369-493-436
6	991	69.3	287	9	US-09-741-669-337
7	991	69.3	287	14	US-10-369-493-23613
8	904.5	63.3	286	14	US-10-369-493-13845
9	891	62.4	289	14	US-10-369-493-8843
10	885.5	62.0	294	14	US-10-369-493-19599
11	838	58.6	287	14	US-10-369-493-15422
12	838	58.6	287	14	US-10-369-493-15790
13	838	58.6	287	14	US-10-369-493-16172

14	812.5	56.9	292	14	US-10-369-493-4487	Sequence 4487, Ap
15	812.5	56.9	292	14	US-10-369-493-7246	Sequence 7246, Ap
16	809	56.6	287	14	US-10-369-493-9371	Sequence 9371, Ap
17	795	55.6	287	14	US-10-369-493-17565	Sequence 17565, A
18	787	55.1	287	14	US-10-369-493-9279	Sequence 9279, Ap
19	585.5	41.0	281	14	US-10-369-493-9902	Sequence 9902, Ap
20	567.5	39.7	293	14	US-10-369-493-8913	Sequence 8913, Ap
21	536.5	37.5	285	14	US-10-369-493-17454	Sequence 17454, A
22	534.5	37.4	315	14	US-10-369-493-20262	Sequence 20262, A
23	533.5	37.3	287	14	US-10-369-493-23329	Sequence 23329, A
24	529	37.0	286	14	US-10-369-493-16552	Sequence 16552, A
25	524.5	36.7	289	14	US-10-310-630-4	Sequence 4, Appli
26	523.5	36.6	289	14	US-10-369-493-18487	Sequence 18487, A
27	509.5	35.7	315	14	US-10-369-493-18941	Sequence 18941, A
28	504.5	35.3	289	14	US-10-310-630-8	Sequence 8, Appli
29	504.5	35.3	314	14	US-10-369-493-2612	Sequence 2612, Ap
30	502	35.1	294	14	US-10-369-493-19549	Sequence 19549, A
31	501.5	35.1	292	16	US-10-474-776-690	Sequence 690, App
32	481	33.7	315	14	US-10-369-493-20871	Sequence 20871, A
33	469.5	32.9	290	14	US-10-369-493-10691	Sequence 10691, A
34	469	32.8	360	15	US-10-767-701-47119	Sequence 47119, A
35	466.5	32.6	301	15	US-10-335-977-5746	Sequence 5746, Ap
36	463	32.4	357	15	US-10-425-114-67588	Sequence 67588, A
37	463	32.4	359	17	US-10-425-115-360784	Sequence 360784,
38	461.5	32.3	325	9	US-09-738-626-4841	Sequence 4841, Ap
39	461.5	32.3	325	16	US-10-781-014-776	Sequence 776, App
40	460.5	32.2	290	14	US-10-369-493-10114	Sequence 10114, A
41	460.5	32.2	293	14	US-10-310-630-12	Sequence 12, Appl
42	459.5	32.2	292	14	US-10-369-493-11433	Sequence 11433, A
43	459.5	32.2	292	14	US-10-369-493-14233	Sequence 14233, A
44	459.5	32.2	292	14	US-10-369-493-14452	Sequence 14452, A
45	459.5	32.2	292	14	US-10-369-493-14863	Sequence 14863, A

## ALIGNMENTS

RESULT 1  
US-09-809-665A-4  
; Sequence 4, Application US/09809665A  
; Publication No. US20040110268A1  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809,665A  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 05/545,199  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 289  
; TYPE: PPT  
; ORGANISM: Pasteurella multocida  
US-09-809-665A-4

Query Match	100.0%	Score 1429;	DB 11;	Length 289;
Best Local Similarity	100.0%;	Pred. No. 2.1e-118;		
Matches 289;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAGAKEIRTKIASVKSTQKI	TKAMWVAASKMKTKQERMSSSRPYSETIRNVISHVSKAT	60
Db	1	MAGAKEIRTKIASVKSTQKI	TKAMWVAASKMKTKQERMSSSRPYSETIRNVISHVSKAT	60
QY	61	IGYKHPFLVDREVKVGMIVVSTDRGLCGGLNVNFKTVLNMKEWKEKDVQSLIGS	120	
Db	61	IGYKHPFLVDREVKVGMIVVSTDRGLCGGLNVNFKTVLNMKEWKEKDVQSLIGS	120	







```

QY 121 KGINFOSLGKILTQDSGIGDTPSVEQLIGVNSMIDAYKKGVDDVVVLYVYNKFTNTMS 180
Db 121 KGVSPFNSVGGNVAAVQVTGMGDNPSLSLIGPVKVMQAVDEGRIDKLYIVSNKFTNTMS 180
QY 181 QKPVLEKLIPLPELDNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240
Db 181 QVPTISQLPLPASDDDL--KHKSWDYLYEPDPKALLDILLARYVBSQVYQGVWENLAS 238
QY 241 EQAARWAMKAATDNAGNLINELQVLYNKARQASITNELNEIVAGAAAI 289
Db 239 EQAARWAMKAATDNGSLIEXLQVLYNKARQASITQELTEIVSGAAAV 287

RESULT 8
US-10-369-493-13845
; Sequence 13845, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13845
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13845

```

```

Query Match 63.3%; Score 904.5; DB 14; Length 286;
Best Local Similarity 61.2%; Pred. No. 7.5e-72;
Matches 177; Conservative 49; Mismatches 60; Indels 3; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASKRKTKQRMSSSRPYSETIRNVISHVSKAT 60
Db 1 MAGAKEIRSKIASIKSTQKITSAEMKAVAKRKAQWRMAASRPYAEIRIQVIGHLANAN 60
QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLPKTVLNEMKWKEDVSVQLSLIGS 120
Db 61 PEYRHPFLVDREVGVVGVVSSDRGLCGGLNLPKTVLNEMKWKEDVSVQLSLIGS 120
QY 121 KGINFOSLGKILTQDSGIGDTPSVEQLIGVNSMIDAYKKGVDDVVVLYVYNKFTNTMS 180
Db 121 KGAAPFFNFNGVNVAAISHLGEPSINDLIGSVKVMQAVDGRIDRLSVSNKFTNTMT 180
QY 181 QKPVLEKLIPLPELDNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENLAS 240
Db 181 QCPVEQILPLVATPDDEL---KHHWDYLYEPDAKELLDMVRYVESQVYQAAVVENNA 237
QY 241 EQAARWAMKAATDNAGNLINELQVLYNKARQASITNELNEIVAGAAAI 289
Db 238 EQAARWAMKAATDNAGDLISDLQLIYNKARQAAITQETSEIVSGAAAV 286

RESULT 9
US-10-369-493-8843
; Sequence 8843, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

```

```

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8843
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Ralstonia metallidurans
US-10-369-493-8843

```

```

Query Match 62.4%; Score 891; DB 14; Length 289;
Best Local Similarity 59.2%; Pred. No. 1.2e-70;
Matches 171; Conservative 57; Mismatches 59; Indels 2; Gaps 1;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASKRKTKQRMSSSRPYSETIRNVISHVSKAT 60
Db 1 MAGTKEIRTKIKSVQNTKTKAMEMVAASKRKKAQERMNARPYAEKVRNIAAHLASAN 60
QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLPKTVLNEMKWKEDVSVQLSLIGS 120
Db 61 PEYKHPFLVDREVKKVGMIVVITDKLGGGLNLPKTVLNEMKWKEDVSVQLSLIGS 120
QY 121 KGINFOSLGKILTQDSGIGDTPSVEQLIGVNSMIDAYKKGVDDVVVLYVYNKFTNTMS 180
Db 121 KGMQFLGRIGAKVSVHVQVQDTPHLEKLGIAIKVQLDAYTNGEVDVAVLYATKINTMK 180
QY 181 QKPVLEKLIPL---PELDNDDELGERKQVWDYIYEPDAKVLDDNLLVRYLESQVYQAAVENL 238
Db 181 QEPWVEQLPLAADKLQSOTDEKRAYSDYIYEPDAQTVVEELLVRYVEALVYQAAENM 240
QY 239 ASEQARWAMKAATDNAGNLINELQVLYNKARQASITNELNEIVAGAA 287
Db 241 ASEQARWAMKAASDNAXNIGELQVLYNKTRQAAITKELSEIVSGAA 289

```

```

RESULT 10
US-10-369-493-19599
; Sequence 19599, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19599
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Nitrosomonas europaea
US-10-369-493-19599

```

```

Query Match 62.0%; Score 885.5; DB 14; Length 294;
Best Local Similarity 56.1%; Pred. No. 3.8e-70;
Matches 165; Conservative 67; Mismatches 57; Indels 5; Gaps 2;

QY 1 MAGAKEIRTKIASVKSTQKITKAMEMVAASKRKTKQRMSSSRPYSETIRNVISHVSKAT 60
Db 1 MPSSREIRNKIKSVQNTKTKAMEMVAASKRKKAQERMNARPYAEKVRNIAAHLASN 60
QY 61 IGYKHPFLVDREVKKVGMIVVSTDRGLCGGLNVNLPKTVLNEMKWKEDVSVQLSLIG 119

```

Db 61 VEYRHPFLSRDSVKRGIIIVTSDKGLCGGLNTNVLRRALNEIRTWETEGNHVDAACIG 120  
QY 120 SKSINFFQSLGKILTDQSGIGDTPSVEQLIGSVNSMIDAYKGEVDVYVLYVYNKFNITM 179  
Db 121 NKGLGPMBSLGTQVTSQVTLGSDAPNMELIGAVKVIDATBQGLDRVYFYFNRFNITM 180  
QY 180 SKQPVLEKLIPIPE-----LDNDELGERKQVMDYIYEPDAKVLNLLVRYLESQVQAAY 235  
Db 181 KQMPVNEQLPLTDDRISSDGEARPTAPMDYIYEPKPVDDIMVRYIEALVYQAVA 240  
QY 236 ENLASQARPMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAI 289  
Db 241 ENVASEQSRPMVAMKAASDNAGNLDELTLVYNKSRQAATQEI SEIVSGAAV 294

RESULT 11  
US-10-369-493-15422  
; Sequence 15422, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 15422  
; LENGTH: 287  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-15422

Query Match 58.6%; Score 838; DB 14; Length 287;  
Best Local Similarity 53.9%; Pred. No. 6.1e-66;  
Matches 158; Conservative 69; Mismatches 56; Indels 10; Gaps 3;  
QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Db 1 MAGGREIKTKIKSVQNTKVTTRALEMVSASKIRKAQERMKTSRPAQAMQVIGHLAQAS 60  
QY 61 IGYKHPFLVDRE-VKKVGMIVVSTDRGLCGGLNVNLFKTVLNEKWEKDVSVQSLIG 119  
Db 61 TDYQHPFLVEREQVRGVYVISSDRGLAGLNNLFRKMLGEVRPMQDKGAEIDVVVTIG 120  
QY 120 SKSINFFQSLGKILTDQSGIGDTPSVEQLIGSVNSMIDAYKGEVDVYVLYVYNKFNITM 179  
Db 121 OKASAFFRIKVMVGSVTHLGDSPHIEQLVGVIKWMLDAFTGKVDVRYLVVNRFNITM 180  
QY 180 SKQPVLEKLIPIPELDNDELGERKQV---WDYIYEPDAKVLNLLVRYLESQVQAAY 236  
Db 181 TOKASFQELLPLP-----AAEHKVAHHDWDYIYEPDAATVLEHVMTRYTESLVYQAVLE 234  
QY 237 NLASEQARPMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAI 289  
Db 235 NVASEHAARPMVAMKAASDNANKMIGTLQVYNKARQAATQEI SEIVSGAAV 287

RESULT 12  
US-10-369-493-15790  
; Sequence 15790, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 15790  
; LENGTH: 287  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-15790

Query Match 58.6%; Score 838; DB 14; Length 287;  
Best Local Similarity 53.9%; Pred. No. 6.1e-66;  
Matches 158; Conservative 69; Mismatches 56; Indels 10; Gaps 3;  
QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Db 1 MAGGREIKTKIKSVQNTKVTTRALEMVSASKIRKAQERMKTSRPAQAMQVIGHLAQAS 60  
QY 61 IGYKHPFLVDRE-VKKVGMIVVSTDRGLCGGLNVNLFKTVLNEKWEKDVSVQSLIG 119  
Db 61 TDYQHPFLVEREQVRGVYVISSDRGLAGLNNLFRKMLGEVRPMQDKGAEIDVVVTIG 120  
QY 120 SKSINFFQSLGKILTDQSGIGDTPSVEQLIGSVNSMIDAYKGEVDVYVLYVYNKFNITM 179  
Db 121 OKASAFFRIKVMVGSVTHLGDSPHIEQLVGVIKWMLDAFTGKVDVRYLVVNRFNITM 180  
QY 180 SKQPVLEKLIPIPELDNDELGERKQV---WDYIYEPDAKVLNLLVRYLESQVQAAY 236  
Db 181 TOKASFQELLPLP-----AAEHKVAHHDWDYIYEPDAATVLEHVMTRYTESLVYQAVLE 234  
QY 237 NLASEQARPMVAMKAATDNAGLNELQVYNKARQASITNELNEIVAGAAI 289  
Db 235 NVASEHAARPMVAMKAASDNANKMIGTLQVYNKARQAATQEI SEIVSGAAV 287

RESULT 13  
US-10-369-493-16172  
; Sequence 16172, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16172  
; LENGTH: 287  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-16172

Query Match 58.6%; Score 838; DB 14; Length 287;  
Best Local Similarity 53.9%; Pred. No. 6.1e-66;  
Matches 158; Conservative 69; Mismatches 56; Indels 10; Gaps 3;  
QY 1 MAGAKEIRTKIASVKSTOKITKAMEMVAASQKRTQERMSSSRPYSETIRNVISHVSKAT 60  
Db 1 MAGGREIKTKIKSVQNTKVTTRALEMVSASKIRKAQERMKTSRPAQAMQVIGHLAQAS 60  
QY 61 IGYKHPFLVDRE-VKKVGMIVVSTDRGLCGGLNVNLFKTVLNEKWEKDVSVQSLIG 119

```

Db      61 TDYQHPFLVEREQKRGYIVISSDRGLAGLNNLFRKMLGEVRFQDKGAIDVVIIG 120
QY      120 SKSINFFQSLGKILTDGSGIGDTPSVEQLIGSVNSMIDAYKGEVDVVYLVYFNKFINTM 179
Db      121 QKASAFFRIRKVMVSGVTHLGDSPHIEQLGVGVKVMKDAFTGKVDVRYLVYFNKFINTM 180
QY      180 SQKPVLEKLIPLPELDNDELGERKQV---WDYIYEPDAKVLNLDLVRYLESQVYAAVE 236
Db      181 TQKASFEQLPLP-----AAEHKVAHHDNDYIYEPDAVLEHVTWTRYIESLVYQAVLE 234
QY      237 NLASEQAAARMVAMKAATDNAGNLINELQVYVYKARQASITNELNEIVAGAAAI 289
Db      235 NVASEAARMVAMKAASDNANKMIGTLQVYVYKARQAAITQEISEIVSGAAAV 287

```

```

RESULT 14
US-10-369-493-4487
; Sequence 4487, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4487
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-369-493-4487

```

```

Query Match      56.9%; Score 812.5; DB 14; Length 292;
Best Local Similarity 53.4%; Pred. No. 1.1e-63;
Matches 158; Conservative 63; Mismatches 62; Indels 13; Gaps 3;

QY      1 MAGAKEIRTKIASYKSTQKITKAMEMVAASKMRKTOERMSSSRPYSETIRNVISHVSKAT 60
Db      1 MAGWKEIRGKIKSVQNTKTKITKAMEMVAASKMRKTOERMSSSRPYSETIRNVISHVSKAT 60
QY      61 IGYKHPFLVDRE-VKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKOVSVQLSLIG 119
Db      61 PEYRHPFMVSNEGAKTAGIILVTTDKLGGGMNTNVLRSLOKFKELGQGGKTIETATAIG 120
QY      120 SKSINFFQSLGKILTDGSGIGDTPSVEQLIGSVNSMIDAYKGEVDVVYLVYFNKFINTM 179
Db      121 TKGLGFLNLRKAVSVNVHLGDTPHLEKLGAVKQVLDLYSEGKVSAYVLYAYTRFVNTM 180
QY      180 SQKPVLEKLIPLPELDNDELGERKQ-----WDYIYEPDAKVLNLDLVRYLESQVY 231
Db      181 KOBEVIEQLPL-----SADQFERKEEDGTTPTQWDYIYEPDAQAVVDLIVRYVEALVY 236
QY      232 QAAVENLASQAAARMVAMKAATDNAGNLINELQVYVYKARQASITNELNEIVAGAA 287
Db      237 QAAVENWASEQAAARMVAMKAASDNAKTVINELQVYVYKARQAAITKELSEIVGGAA 292

```

```

RESULT 15
US-10-369-493-7246
; Sequence 7246, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

```

```

; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7246
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-369-493-7246

```

```

Query Match      56.9%; Score 812.5; DB 14; Length 292;
Best Local Similarity 53.4%; Pred. No. 1.1e-63;
Matches 158; Conservative 63; Mismatches 62; Indels 13; Gaps 3;

QY      1 MAGAKEIRTKIASYKSTQKITKAMEMVAASKMRKTOERMSSSRPYSETIRNVISHVSKAT 60
Db      1 MAGWKEIRGKIKSVQNTKTKITKAMEMVAASKMRKTOERMSSSRPYSETIRNVISHVSKAT 60
QY      61 IGYKHPFLVDRE-VKKVGMIVVSTDRGLCGGLNVLNFKTVLNEMKEWKEKOVSVQLSLIG 119
Db      61 PEYRHPFMVSNEGAKTAGIILVTTDKLGGGMNTNVLRSLOKFKELGQGGKTIETATAIG 120
QY      120 SKSINFFQSLGKILTDGSGIGDTPSVEQLIGSVNSMIDAYKGEVDVVYLVYFNKFINTM 179
Db      121 TKGLGFLNLRKAVSVNVHLGDTPHLEKLGAVKQVLDLYSEGKVSAYVLYAYTRFVNTM 180
QY      180 SQKPVLEKLIPLPELDNDELGERKQ-----WDYIYEPDAKVLNLDLVRYLESQVY 231
Db      181 KOBEVIEQLPL-----SADQFERKEEDGTTPTQWDYIYEPDAQAVVDLIVRYVEALVY 236
QY      232 QAAVENLASQAAARMVAMKAATDNAGNLINELQVYVYKARQASITNELNEIVAGAA 287
Db      237 QAAVENWASEQAAARMVAMKAASDNAKTVINELQVYVYKARQAAITKELSEIVGGAA 292

```

Search completed: November 9, 2004, 10:00:08  
Job time : 141 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 9, 2004, 05:20:38 ; Search time 23949 Seconds  
(without alignments)  
570.659 Million cell updates/sec

Title: US-09-545-199F-4

Perfect score: 1429

Sequence: 1 MAGAKEIRTKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USFO\_spool/US09545199/runat\_08112004\_112526\_20949/app\_query.fasta\_1.455  
-DB=genembl -Qfmt=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -NATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09545199 @CGN 1 1 7406 @runat\_08112004\_112526\_20949 -NCFU=6 -ICPU=3  
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.btg.\*  
3: gb.in.\*  
4: gb.on.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1429	100.0	870	1	AF237922 Pasteurel
2	1429	100.0	1972	6	BD252118 Anti-bact
3	1429	100.0	1972	6	AX553669 Sequence
4	1429	100.0	10229	1	AE006185 Pasteurel

C	5	1173	82.1	13828	1	U32730
C	6	1173	82.1	110000	6	AR274513_04
C	7	1173	82.1	110000	6	AR274513_05
C	8	1132.5	79.3	867	6	BD252190
C	9	1132.5	79.3	867	6	AX553798
C	10	1121.5	78.5	304558	1	AE017151
C	11	1077.5	75.4	866	6	AX553832
C	12	1004	70.3	876	6	AX378271
C	13	1004	70.3	349483	1	BX571859
C	14	988.5	69.9	12928	1	AE004342
C	15	993	69.5	64019	1	EX950851_50
C	16	991	69.3	864	6	AX188943
C	17	991	69.3	2835	1	ECATPY
C	18	991	69.3	7141	1	ECOHATP
C	19	991	69.3	7421	1	AF188265
C	20	991	69.3	11999	1	AE005605
C	21	991	69.3	12599	1	AE015388
C	22	991	69.3	14526	1	ECUNC
C	23	991	69.3	110000	1	U00096_39
C	24	991	69.3	136254	1	ECCJW82
C	25	991	69.3	280900	1	AP002566
C	26	991	69.3	293558	1	AE016991
C	27	991	69.3	301566	1	AE016769
C	28	989	69.2	11099	1	AE014015
C	29	989	69.2	203728	1	AJ414160
C	30	989	69.2	229338	1	AE017142
C	31	988	69.1	7881	1	ECUNC
C	32	988	69.1	20518	1	AE008880
C	33	982.5	68.8	10718	1	AE015907
C	34	980.5	68.6	104755	1	AP005343
C	35	980.5	68.6	299938	1	AE016800
C	36	978	68.4	894	6	AR386716
C	37	978	68.4	245050	1	AL627280
C	38	978	68.4	300592	1	AE016846
C	39	976.5	68.3	7402	1	VAUNC
C	40	970.5	67.9	288108	1	AP005083
C	41	947.5	66.3	251097	1	CR378674
C	42	931.5	65.2	7863	1	AB035129
C	43	926.5	64.8	16662	1	AE004967
C	44	915	64.0	349544	1	CR378675
C	45	896.5	62.7	142665	1	AE016794

#### ALIGNMENTS

AF237922	Pasteurella multocida ATP synthase F1 gamma chain (atpg) Gene, complete cds.	870 bp	DNA	linear	BCT 08-MAY-2000
AF237922	AF237922.1	GI:7716504			
Pasteurella multocida					
Pasteurella multocida					
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;					
Pasteurellaceae; Pasteurella.					
1 (bases 1 to 870)					
Fuller,T.E., Kennedy,M.J. and Lowery,D.E.					
Identification of Pasteurella multocida virulence genes in a septicemic mouse model using signature-tagged mutagenesis					
Unpublished					
2 (bases 1 to 870)					
Fuller,T.E., Kennedy,M.J. and Lowery,D.E.					
Direct Submission					
Submitted (24-FEB-2000) Discovery Research, Pharmacia & Upjohn					
Animal Health, 7923-25-434, 7000 Portage Road, Kalamazoo, MI					
49001-0199, USA					
Location/Qualifiers					
1. .870					
/organism="Pasteurella multocida"					
/mol_type="genomic DNA"					
/db_xref="taxon:747"					

```

gene      1..870
          /gene="atpg"
CDS       1..870
          /gene="atpg"
          /codon_start=1
          /transl_table=11
          /product="ATP synthase F1 gamma chain"
          /protein_id="AAF68408.1"
          /db_xref="GI:7715505"
          /translation="MAGAKETIKASVSTQKITKAMENVAASKVKRKTQERMSSRP
YSETRINVISHSKATIGYKHPFLVREKVGMIIVSTDRGLCGGLNVLNFKTVLNE
MKEWKEVSVQLSLIGSKINFPQSLGKILIQDSGIGDTPSEVLGSLGVNSMTDAY
KKGVDVYLVKINFMQKPVLEKLIPLPELDNDELGERKQVWDXYIERDAKVL
DNLIVLYESQVYAAVENLASQAARWAMKATDNAGNLINELQLVYVKARQASIT
NELNEIVAGAAI"

ORIGIN
Alignment Scores:
Pred. No.: 1,48-110 Length: 870
Score: 1429.00 Matches: 289
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-545-199F-4 (1-289) x AAF237922 (1-870)
QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
DB 1 ATGCAGCGTCTAAAGAGATAAGAACCAAAATCGCAGTGTAAAGATACACAAAAAATT 60
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
DB 61 ACTAAAGCGATGAAATGGTGTGCTCGAAATCGTAAACCAAGCAAGCATGTCT 120
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
DB 121 TCTTCAGCCCTTATTCAGAACCAATACGTACGTGATTTAGCCAGTTTCCAAAGCAACG 180
QY 61 IleGlyTyrLysHisProPheLeuValAlaAspArgGluValLysValGlyMetIleVal 80
DB 181 ATTGGTTTACAAGATCCATTTTATGTGGATCGGAAGTAAAAAAGTGGCATGATGTT 240
QY 81 ValSerThrAspArgGlyLeuCysGlyLeuAsnValAsnLysPheLysThrValLeu 100
DB 241 GTGTCCACAGATCGTGCTTTGTGTGGCTTAAACGTGAATCTGTTTAAACATGTATTA 300
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
DB 301 AATGAAATGAAGATGGAAGAAAAAGATGTTCCGTTCAATTGAGTTTAAATCGGTTCT 360
QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
DB 361 AAATCTATCACTTTTCCAACTTTGGGAATTAATAATTTAAACCAAGATTCAGGTATT 420
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
DB 421 GGTGATACTCCCTCTGTCAGCAGTAAATTTGTTCCAGTCAATTTATGATGATGCTTAT 480
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
DB 481 AAAAAAGGGAGATGATGTTGTGATTTAGTTTATATAACAAATTTATTAACAGATGTCG 540
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
DB 541 CAAAGCCAGTATTGAAAAATTAATTCATTAACAGATTTAGATATGATGATTTAGGC 600
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGlnProAspAlaLysValLeuLeuAspAsn 220
DB 601 GAAAGAAACAAGTTTGGATTTATTTTACGAACCTGATCGAAAGTATTATTAGATAAT 660
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240

```

```

Db 661 TTATTGGTTCGTTATTAGAACTCTCAGGTTTATCAAGCAGCAGTTTCAAAACCTTCTTCT 720
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
DB 721 GAGCAGCCGCTCGAATGGTTCGCCATGAAAGCAGCAACAGATAACGAGTAACATAATT 780
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
DB 781 AATGAGTTACAGTTAGTCTATACAAAGCTCTCAAGCAAGTATTACAAATGAATTAAT 840
QY 281 GluIleValAlaGlyAlaAlaIle 289
DB 841 GAAATGTTGCGCGTCGACAGCAAT 867

RESULT 2
BD252118 1972 bp DNA linear PAT 17-JUL-2003
LOCUS Anti-bacterial vaccine compositions.
DEFINITION BD252118
ACCESSION BD252118
VERSION 1.1
KEYWORDS GI:33061888
SOURCE JP 2002541790-A/2.
ORGANISM Pasteurella multocida
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Pasteurella.
REFERENCE 1 (bases 1 to 1972)
AUTHORS Lowery,D.E., Fuller,T.E. and Kennedy,M.J.
TITLE Anti-bacterial vaccine compositions
JOURNAL Patent: JP 2002541790-A 2 10-DEC-2002;
PHARMACIA AND UPJOHN CO
COMMENT OS Pasteurella multocida
PN JP 2002541790-A/2
PD 10-DEC-2002
PP 06-APR-2000 JP 2000611649
PR 09-APR-1999 US 60/128689,10-SEP-1999 US 60/153453 PI
DAVID E LOWERY, TROY E FULLER, MICHAEL J KENNEDY PC
C12N15/09,A61K39/02,A61K39/102,A61K39/395,A61P31/04,C07K14/285, PC
C07K16/12,
PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N15/01,C12P21/02 PC
C12P21/08,C12Q1/18,
PC G01N33/15,G01N33/50,G01N33/53,G01N33/569,G01N33/577,C12N15/00,
PC C12N15/00,
PC C12N5/00
CC atpg
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..1972 (364)..(1230).
/organism="Pasteurella multocida"
/db_xref="taxon:747"

ORIGIN
Alignment Scores:
Pred. No.: 3,85e-110 Length: 1972
Score: 1429.00 Matches: 289
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-545-199F-4 (1-289) x BD252118 (1-1972)
QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
DB 364 ATGCAGCGTCTAAAGAGATAAGAACCAAAATCGCAGTGTAAAGATACACAAAAAATT 423
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
DB 424 ACTAAAGCGATGGAATGTTGTGCTCGAAATCGTAAACCAAGCAAGCATGTCT 483
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60

```

```
Db 484 TCTTCAGCCCTTATTTCAGAAACAATACGTAACTGATAGCCACGTTTCCAAAGCAACG 543
Qy 61 ILeGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Db 544 ATTGGTTTACAAGCATCCATTTTGTAGTGATCGCGAAGTAAATAAAGTGGCATGATTGT 603
Qy 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAenValAsnLeuPheLysThrValLeu 100
Db 604 GTCTCACAGATCGTGGTCTTTGTGTGGCTTAAACGTAACCTTGTATTAACATGTTATTA 663
Qy 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db 664 AATGAATGAAGAATGGAAGAAAGATGTTTCGGTTCATTCAGTTTAAATCGGTCTCT 723
Qy 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db 724 AATCTATCACTTTTCCAAATCTTTGGGAATTAATAATTTTAAACCAAGATTCAGGTATT 783
Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 784 GGTGATACCTCCCTCTGTGTAGCAGTAAATTTAGTGTTCAGTCAATTCATGATTCGCTTAT 843
Qy 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 844 AAAAAAGGGGAAGTAGATCTTGTGATTTAGTTTATAACAAATTTATTAACACGATGTCG 903
Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 904 CAAAGCCAGTATTGGAAAAATTAATTCATTCACGAATATTAGATATGATGAATTAGGC 963
Qy 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 964 GAAAGAAAAAAGTTGGGATATATTTACGAACCTGATCGAAAGTATTATTAGATAAT 1023
Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaLysValLeuLeuAspAsn 240
Db 1024 TTATGTGTTGCTTATTAGATCTCAGGTTTATCAAGCAGCAGTGTGAAACCTTGCTTCT 1083
Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db 1084 GAGCAAGCCGCTCGAATGTCGCCATGAAGCAGCAACAGATACGCGAGTAACTTAAT 1143
Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db 1144 AATGAGTTACAGTTAGTCTTATAACAAAGCTCGTCAAGCAAGTATTACAAATGAATTAAT 1203
Qy 281 GluIleValAlaClyAlaAlaIle 289
Db 1204 GAAATGTGCGGTCGACGACCAATT 1230

RESULT 3
AX553669 LOCUS 1972 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 3 from Patent WO02075507.
ACCESSION AX553669
VERSION AX553669.1 GI:25897667
KEYWORDS Pasteurella multocida
SOURCE Pasteurella multocida
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Pasteurella.
REFERENCE 1
AUTHORS Lowery, D.E., Fuller, T.E. and Kennedy, M.J.
TITLE Anti-bacterial vaccine compositions
JOURNAL Patent: WO 02075507-A 3 26-SEP-2002;
Pharmacia & Upjohn Company (US)
FEATURES
    Location/Qualifiers
        1..1972
            /organism="Pasteurella multocida"
            /mol_type="unassigned DNA"
            /db_xref="taxon:747"
            364..1233
            /gene="atpG"
```

## CDS

```
364..1233
/gene="atpG"
/codon_start=1
/transl_table=11
/protein_id="CAD58126.1"
/db_xref="GI:25897668"
/translation="MAGAKIRTKIASVSKTKITKAMVMVAASKKRKTQERSSSRP
YSETRNVISHVSKATIGYKHPFLVDREVKKVMIVVSTDRGLCGGLNVNLFKTVLNE
MKEMKEDVSVQLASLIGSKSINFFQSIGIKLITQDSIGIGTSPSEQILGVSNSKIDAY
KKGEVDVYLYNKFINTMSKPLVKLILPLDNDDELGERKQWDYIYEPDAKVLL
DNLVRYLESQYQRAVENLASEQARVAMKAATDAGNLINELQLVYNKARQASIT
NELNEIVAGAAAI"
```

## ORIGIN

## Alignment Scores:

```
Pred. No.: 3,85e-110 Length: 1972
Score: 1429.00 Matches: 289
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
```

US-09-545-199F-4 (1-289) x AX553669 (1-1972)

```
Qy 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
Db 364 ATGGCAGGTGCTAAAGAGATAGAACCAAAATCCGAGTGTAAATAAGTACACAAAAAATT 423
Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
Db 424 ACTAAAGCGATGGAATGGTGTGCTCGAATGCGTAAACGCGAAGACGATGTCT 483
Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db 484 TCTTCACGCCCTTATTAGTGTGCTTGTGGTGGCTTAAACGCGTAAACCTGTTTAAACCTGATT 603
Qy 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Db 544 ATTGGTTACAAGCATCCATTTTGTAGTGATCGCGAAGTAAATAAAGTGGCATGATTGT 603
Qy 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db 604 GTGTCCACAGATCGTGGTCTTGTGTGGTGGCTTAAACGCGTAAACCTGTTTAAACCTGATT 663
Qy 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db 664 AATGAATGAAGAATGGAAGAAAGATGTTTCGGTTCATTCAGTTTAAATCGGTCTCT 723
Qy 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db 724 AATCTATCACTTTTCCAAATCTTTGGGAATTAATAATTTTAAACCAAGATTCAGGTATT 783
Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 784 GGTGATACCTCCCTCTGTGTAGCAGTAAATTTAGTGTTCAGTCAATTCATGATTCGCTTAT 843
Qy 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 844 AAAAAAGGGGAAGTAGATCTTGTGATTTAGTTTATAACAAATTTATTAACACGATGTCG 903
Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 904 CAAAGCCAGTATTGGAAAAATTAATTCATTCACGAATATTAGATATGATGAATTAGGC 963
Qy 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 964 GAAAGAAAAAAGTTGGGATATATTTACGAACCTGATCGAAAGTATTATTAGATAAT 1023
Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaLysValLeuLeuAspAsn 240
Db 1024 TTATGTGTTGCTTATTAGATCTCAGGTTTATCAAGCAGCAGTGTGAAACCTTGCTTCT 1083
Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
```

```

Db      1084 GAGCAAGCGCTCGAATGCTGCCATGAAGACGACGACAGATAACGACAGTAACTTAATT 1143
QY      261 AenGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db      1144 AATGAGTTACAGTATAGTCTTATACAAAGCTCGTCAAGCAAGTATTACAAATGAATTAAT 1203
QY      281 GhuLeuValAlaGlyAlaAlaLeu 289
Db      1204 GAAATGTTGCGGTGCAGCAACAATT 1230

RESULT 4
LOCUS   AE006185                10229 bp    DNA        linear    BCT 16-JUN-2004
DEFINITION Pasteurella multocida subsp. multocida str. Pm70 section 152 of 204
of the complete genome.
ACCESSION AE006185 AE004439
VERSION   AE006185.1 GI:12721868
KEYWORDS
SOURCE   Pasteurella multocida subsp. multocida str. Pm70
ORGANISM Pasteurella multocida subsp. multocida str. Pm70
          Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
          Pasteurellaceae; Pasteurella.
REFERENCE 1 (bases 1 to 10229)
AUTHORS   May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
          Kapur,V.
TITLE      Complete genomic sequence of Pasteurella multocida, Pm70
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE   21145866
PUBMED    11248100
REFERENCE 2 (bases 1 to 10229)
AUTHORS   Zhang,Q. and Kapur,V.
TITLE      Direct Submission
JOURNAL   Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
          University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
          55108, USA

FEATURES             Location/Qualifiers
     source           1..10229
                     /organism="Pasteurella multocida subsp. multocida str.
                     Pm70"
     mol_type="genomic DNA"
     strain="PM70"
     sub_species="multocida"
     db_xref="taxon:272843"
     275..2164
     /gene="gida"
     /note="synonym: PM1485"
     275..2164
     /gene="gida"
     /codon_start=1
     /transl_table=11
     /product="GidA"
     /protein_id="AAK03569.1"
     /db_xref="GI:12721869"
     /translation="MFYTENVDVIVIGGHAGTEAALAPAWGLTKILLTHNVDTLQ
     MSCNPAIGIGKGLHVRIDAMGVMATADQAGIQFTLNSSKGPVAVRATRAQDRV
     LVQAVRIALENCNLDIQOEVTBIIIDQVCGVVTWGMGLKFAKAVILTAGTFLS
     GKIHGLEYNTGAGDPASVMDLRLEMLNRVDRKTGTPPRIDARTIDFSLAQ
     HGEKLPVFSMGSDVQHPQIPCFITTHNEQTHEIVIRNNLDSPMTAGIIEGIGPRY
     CPSIEDKVRFSERNSHQIYLEPGLTSNIEYPNGISLSPFDVQMKIVMSKMGKA
     RIIPGYAIETDYFDPRDLKPTLETKTSIRGLFPAQVINGTGYEEAAGQGLLAGINAG
     LFVQEKAAFPRRDQAYIGLVLDLCTLTGKTPVRVTSRAEYELLIRENADSRITP
     IAHOLGIDIKERWARFNQWENIELEORLQIWLHRSVYLDENKVLGSPFLVRAS
     GDLLRRPMYQILTSITPFPAMDQEAQEVEIAIKYQIEHQOEIEIAQKRHE
     STAPFAHYTYTVVSGLSNEVRKAEQHRPVSIGQASRISGVTAAISILVSLKQGM
     LRGE"
     2164..2796
     /gene="gidB"
     /note="synonym: PM1486"
     2164..2796
     /gene="gidB"
     /codon_start=1
     /transl_table=11

```

```

/product="GidB"
/db_xref="GI:12721870"
/translation="MTNLEQQLSQKLEILLKOTALSITDQOEOLIKLVLLANKKKA
YNTSVRDPDMELVKHILDSVVSPHLOQKHFDIVGTGPGELGLPLATVFNPKHFVLL
DSLGRISFTRNAIRBELGLDNVEAVLSRVEEHPQPPDGVLSRAFASLKMDTWCOQH
LPKQDGYVALKGLYHQEEVEELSERFTIQQVIRLQVPELIGERHLVIVK"
2934..3341
/gene="PM1487"
2934..3341
/gene="PM1487"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03571.1"
/db_xref="GI:12721871"
/translation="MSKVIYQAKQKYKALGIELGLVLLASLTCLIGYQWQVAISFILG
SSAIFPHCLFVNFVFTKKQFSNKLTFYGGEAVKFTSTIFLIVLAFKLPNSMHPFI
AFFVGYFITVLNNILPFFISKYLNKHPHYQ"
3352..4146
/gene="atpB"
/note="synonym: PM1488"
3352..4146
/gene="atpB"
/codon_start=1
/transl_table=11
/product="AtpB"
/protein_id="AAK03572.1"
/db_xref="GI:12721872"
/translation="MAAELTTAGYIGHLAFIKTGDSPWVHVDTLPSIISGAIFLF
VSKVAKATPGVPSKQCFVEIMVDWIDGIVKGNFHPGRHAGVGLPALTIFCWYFIMN
ADLPVDFPLQALHFGLEYLRVAPTADISGTGLSGISGVFLIIFTIKSGMSGFV
KSTYHPNHLPIIPNLALSESVTLLAKPVSIAFLFCNMYAGELIFTLIAVMYMANN
PALNSGIFPMHLAWAIFHLITLQAFIFMLTVVLSMGYNKAEH"
4203..4457
/gene="atpE"
/note="synonym: PM1489"
4203..4457
/gene="atpE"
/codon_start=1
/transl_table=11
/product="AtpE"
/protein_id="AAK03573.1"
/db_xref="GI:12721873"
/translation="METVITITIASAILLAVALGATGALFAILGKGFLESSARQPEL
ASSQIKMFIAGLLDAISMAVGIALLFIANPFIDLLK"
4537..5007
/gene="atpF"
/note="synonym: PM1490"
4537..5007
/gene="atpF"
/codon_start=1
/transl_table=11
/product="AtpF"
/protein_id="AAK03574.1"
/db_xref="GI:12721874"
/translation="WNLNATLIGLIAIAFVAFCKFVWPPIKAIEREQSIANAL
ASAAKQEQADTALVEQITLAKMQAQOIIILANKRNEIEEVEKVEAEATKAKII
EQGVAEAEAKRVQBELRVKASLAAGAEKIVGRTVDEANSIDIKLVAEL"
5021..5554
/gene="atpH"
/note="synonym: PM1491"
5021..5554
/gene="atpH"
/codon_start=1
/transl_table=11
/product="AtpH"
/protein_id="AAK03575.1"
/db_xref="GI:12721875"
/translation="MSBELTIAIRPYAKAAAFDAIEQNAVEKWNMLHFSSSLIKDET
QTFKSSLSASKLADTVISICGQDQVGNLIRLMAENKRLVLPALYQAFQHVVEE
HOKVAEQVISAQPLATQKIASAMEKRLARKVKLNCSLDSSLIAGVIIRTDIFVI
DGSSRGQLSRLANELQL"

```

gene	5567..7108	Query Match:	100.00%	Indels:	0
	/gene="atpA"	DB:	1	Gaps:	0
CDS	5567..7108	US-09-545-199f-4 (1-289) x AE006185 (1-10229)			
	/gene="atpA"	QY	1	MetAlaGlyAlaLysGluLeuArgThrLysIleAlaSerValLysSerThrGlnLysIle	20
	/codon_start=1	DB	7135	ATGCGAGGTGCTAAAGAGATAAGAACCAAAATCCGAGTCTAAAGAGTACACAAAAAT	7194
	/transl_table=11	QY	21	ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer	40
	/product="atpA"	DB	7195	ACTAAAGCGATGGAAATGGTTCCTCGAAATGCGTAAACCGCAAGACCGATGTCT	7254
	/db_xref="GI:12721876"	QY	41	SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr	60
	/db_xref="AAK03576.1"	DB	7255	TCCTCAGCCCTTATTTCAGAAACAATACGTAACTAGCCACGCTTCCCAAGCAACG	7314
		QY	61	IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal	80
		DB	7315	ATTGGTTACAGCATCCATTTTATGTCGATCGGAAGTAAAGATGGGCATGATTGT	7374
gene	7135..8004	QY	81	ValSerThrAspArgGlyLeuCysGlyLeuAsnValAsnLeuPheLysThrValLeu	100
	/gene="atpC"	DB	7375	GTGTCCACAGATCGTGTCTTTGTGTGGCTTAAACGTGAACCTTGTAAACCTGATT	7434
CDS	7135..8004	QY	101	AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer	120
	/codon_start=1	DB	7435	ANTGAATGAAGAAATGGAAGAAAGATGTTCTCTCAATTTGATTTAAATCGTTCT	7494
	/transl_table=11	QY	121	LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle	140
	/product="atpC"	DB	7495	AAATCTATCACTTTTCCCAATCTTTGGGAATTAATAATTTAAACCAAGATTCAGT	7554
	/protein_id="AAK03577.1"	QY	141	GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr	160
	/db_xref="GI:12721877"	DB	7555	GGTGATACTCCCTCTGTGTGACGATTAATTTGGTTCAGTCAATTTCTATGATTGAT	7614
gene	8022..9395	QY	161	LysLysGlyGluValValValValValValValValValValValValValValVal	180
	/gene="atpD"	DB	7615	AAAAAGGGGAGAGTAGATGTTGTGTATTTAGTTTAAACAAATTTATTACACAGAT	7674
CDS	8022..9395	QY	181	GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAspGluLeuGly	200
	/codon_start=1	DB	7675	CAAAAGCCAGTATTGGAATAATTAATTCATTCACAGATTAAGATAATGATGAAT	7734
	/transl_table=11	QY	201	GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuAspAsn	220
	/product="atpD"	DB	7735	GAAAGAAACAGTTTGGGATTTATTTACGAACTGTATCGAAAGTATTATTAGATA	7794
	/db_xref="GI:12721878"	QY	221	LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGlnValGluAsnLeu	240
		DB	7795	TTATTGGTTCTGTTTATTAATCTCAGGTTTATCAAGCAGCAGGTTGAAACCTT	7854
gene	9436..9864	QY	241	GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeu	260
	/gene="atpC"	DB	7855	GAGCAAGCCGCTCGAATGTCGCCATGAAGCAGCAACAGATTAACGAGGTAAT	7914
CDS	9436..9864	QY	261	AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeu	280
	/codon_start=1	DB	7915	AATGAGTTACAGTTAGTCTTATTAACAAAGCTGCTCAAGCAGGATTACAAATGA	7974
	/transl_table=11	QY	281	GluLeuValAlaGlyAlaAlaAlaIle	289
	/product="atpC"	DB	7975	GAAATTTGTCGGGTGCAGCAAT	8001

RESULT 5  
U32730/c  
LOCUS

U32730  
Haemophilus influenzae Rd KW20 section 45 of 163 of the complete genome  
U32730 L42023  
U32730.1 GI:3212191  
KEYWORDS

Alignment Scores:  
Pred. No.: 2,95e-109 Length: 10229  
Score: 1429.00 Matches: 289  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0



SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

Haemophilus influenzae Rd KW20  
Haemophilus influenzae Rd KW20  
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
Pasteurellaceae; Haemophilus.

1 (bases 1 to 13828)  
Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,  
Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,  
McCarthy, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,  
Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,  
Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,  
Cotton, M.D., Utterback, T., Hanna, M.C., Spriggs, D.T., Saudek, D.M.,  
Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Saudek, D.M.,  
Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,  
Smith, H.O. and Venter, J.C.

Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd  
Science 269 (5223), 496-512 (1995)  
95350630  
PUBMED

2 (bases 1 to 13828)  
Taturov, R.L., Mushagian, A.R., Bork, P., Brown, N.P., Hayes, W.S.,  
Borodovsky, M., Rudd, K.E. and Koonin, E.V.

Metabolism and evolution of Haemophilus influenzae deduced from a  
whole-genome comparison with Escherichia coli  
Curr. Biol. 6 (3), 279-291 (1996)  
96398784  
PUBMED

3 (bases 1 to 13828)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.

Direct Submission  
TITLE  
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

4 (bases 1 to 13828)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.

Direct Submission  
TITLE  
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

The H. influenzae sequence has been updated by R. Fleischmann. New  
database matches have been assigned, product names have been  
improved, and a number of frame shifts have been corrected. We  
gratefully acknowledge the work of Taturov et. al. We have  
incorporated their annotation into the /notes fields of the  
corresponding H. influenzae genes

5 (bases 1 to 13828)  
White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,  
Peterson, J., Hickey, E., Dodson, R. and Winn, M.

Direct Submission  
TITLE  
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

The whole genome was shifted by 588 nucleotides for a new start  
On Jun 12, 1998 this sequence version replaced gi:1573448.  
Location/Qualifiers

1..13828  
/organism="Haemophilus influenzae Rd KW20"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:71421"  
65..1168  
/gene="HI0470"  
65..1168  
/gene="HI0470"  
/note="similar to SP:P06986 GB:X03416 PID:41695  
PID:41710 percent identity: 60.40; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="histidinol-phosphate aminotransferase (hisC)"  
/protein\_id="AAC22132.1"  
/db\_xref="GI:1573449"

gene  
CDS

VOALTOELLSNRQWLAKALLVHVEKYVSEANYLLIKQNGOAVFKALWEQGIILR  
DNKLHLQNCIRIIVGRNECEKVEIAKEVK"  
1280..2368  
/gene="HI0471"  
1280..2368  
/gene="HI0471"  
/note="similar to SP:P06987 GB:X03416 PID:41696 PID:41711  
GB:U00096 percent identity: 64.97; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="imidazoleglycerol-phosphate dehydratase /  
histidinol-phosphatase (hisB)"  
/protein\_id="AAC22130.1"  
/db\_xref="GI:1573450"  
/translation="MQPTLFTDRDGLIDEPKTDFOIDLSLEKLEPKVPIALLRLKA  
KYRFVIVSNQDGLGTDAFPQDFKPHVMNALFESQGITFDEVILCFHPKPENCICR  
KPKIKLLDHYIRKNLFDIDRSFVIGDRETDVQLAENLIGIRAIQYDPQKMWDLIAEKL  
LGRTVTCGKPPRPAEVIQRTKEDTIKVQWLDGAGNBIKTGVGFDMQLQIATH  
GGFRNVQCKGDLWIDEHTVEDTALGQALQKQAVGDKRGIGARFGVFLGMDCKARC  
ALDLSGRPWIKFNACFKRDKVGFSTELTEHFFQSLAFSLATLHLNVTGNDHDKIE  
SLFKAFGRTLRQAIRIEGNEPSSKGLV"  
2434..3033  
/gene="HI0472"  
2434..3033  
/gene="HI0472"  
/note="similar to SP:P10375 PID:41712 GB:U00096  
PID:1736701 PID:1736710 percent identity: 55.90;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="amidotransferase (hisH)"  
/protein\_id="AAC22131.1"  
/db\_xref="GI:1573451"  
/translation="MINITITDGCANLSSVFAFRLGYNTBITFDLNNKIKSADKLI  
LNRIPDTGTAAMYLQERLQETIQLNLTQPVGLICGLMQLMTSESGNPTLNLSGK  
TFNRTDLPQPLQPMQNRVQFVKNCPLFDGIQVNSHFYFVHSYAVSFNEHSVAISNG  
VNFSAATAKENFGVQVQHPERSGKNGALLKNFVKVFF"  
3069..3818  
/gene="HI0473"  
3069..3818  
/gene="HI0473"  
/note="similar to SP:P10371 PID:41713 GB:U00096  
PID:1736702 PID:1736711 percent identity: 62.86;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoribosylformimino-5-aminimidazole  
carboxamide ribotide isomerase (hisA)"  
/protein\_id="AAC22132.1"  
/db\_xref="GI:1573452"  
/translation="MKQSIIPALDLINGVVRHOGDYAKOTYSDNPIKQFDNYVR  
QGAQLHLVDLTGAKNPQSRQALIKIVEATQCKVQVGGRTBDQVADLLAVGANR  
VWIGSTATRSVWQWFKYGAERFVLADVNINNASGQKIVAIQSGOEESGVILLET  
IEDPQVGLQQLVCTDISRDQTLTSGNIYQIECKYPIQFQSSGGIGSLADIEAL  
KGTGVSQVIVGRALLEGKFTLSAIAKWQNG"  
3800..4576  
/gene="HI0474"  
3800..4576  
/gene="HI0474"  
/note="similar to SP:P10373 PID:41714 PID:598474 GB:U00096  
PID:1736704 percent identity: 82.03; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hisF cyclase (hisF)"  
/protein\_id="AAC22133.1"  
/db\_xref="GI:1573453"  
/translation="MLAKRILFCLDVRQGVKQVQFRNHEIIGDIVPLAQRVQEGA  
DELVFDITASDGRITVDKSRQVADIVDFCVAGGKITIEDAKLIFAGADKISI  
NSPALADPILRLADRFQVQAIIVGIDSWFEQETGKYWNQYTCDETRTQHWLL  
DWFQVQCCAGIEVLNMNODLGRNGYDLAQLKLVRCVCRVPLIASGAGKMWHRD  
AFTEAKVDGALAAVPHKQIIEIGELKSVLQSAIETRSE"

gene  
CDS

Haemophilus influenzae Rd KW20  
Haemophilus influenzae Rd KW20  
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
Pasteurellaceae; Haemophilus.

1 (bases 1 to 13828)  
Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,  
Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,  
McCarthy, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,  
Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,  
Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,  
Cotton, M.D., Utterback, T., Hanna, M.C., Spriggs, D.T., Saudek, D.M.,  
Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Saudek, D.M.,  
Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,  
Smith, H.O. and Venter, J.C.

Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd  
Science 269 (5223), 496-512 (1995)  
95350630  
PUBMED

2 (bases 1 to 13828)  
Taturov, R.L., Mushagian, A.R., Bork, P., Brown, N.P., Hayes, W.S.,  
Borodovsky, M., Rudd, K.E. and Koonin, E.V.

Metabolism and evolution of Haemophilus influenzae deduced from a  
whole-genome comparison with Escherichia coli  
Curr. Biol. 6 (3), 279-291 (1996)  
96398784  
PUBMED

3 (bases 1 to 13828)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.

Direct Submission  
TITLE  
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

4 (bases 1 to 13828)  
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.

Direct Submission  
TITLE  
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

The H. influenzae sequence has been updated by R. Fleischmann. New  
database matches have been assigned, product names have been  
improved, and a number of frame shifts have been corrected. We  
gratefully acknowledge the work of Taturov et. al. We have  
incorporated their annotation into the /notes fields of the  
corresponding H. influenzae genes

5 (bases 1 to 13828)  
White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,  
Peterson, J., Hickey, E., Dodson, R. and Winn, M.

Direct Submission  
TITLE  
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

The whole genome was shifted by 588 nucleotides for a new start  
On Jun 12, 1998 this sequence version replaced gi:1573448.  
Location/Qualifiers

1..13828  
/organism="Haemophilus influenzae Rd KW20"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:71421"  
65..1168  
/gene="HI0470"  
65..1168  
/gene="HI0470"  
/note="similar to SP:P06986 GB:U02071 GB:X03416 PID:41695  
PID:41710 percent identity: 60.40; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="histidinol-phosphate aminotransferase (hisC)"  
/protein\_id="AAC22132.1"  
/db\_xref="GI:1573449"

gene  
CDS

```

gene      4576..5241
CDS       /gene="HI0475"
          4576..5241
          /gene="HI0475"
          /note="similar to SP:P06989 GB:D43637 GB:U02072 GB:X03974
          PID:41700 percent identity: 61.22; identified by sequence
          similarity/putative"
          /codon_start=1
          /transl_table=11
          /product="phosphoribosyl-AMP cyclohydrolase /
          phosphoribosyl-ATP pyrophosphatase (hisE)"
          /protein_id="AAC22134.1"
          /db_xref="GI:1573454"
          /translation="MLTKIDWQYNGLLPVIQNIISTREVLMGLMNEEALTKTIKE
          KVTFFSRTRQKRWKGEISGNFNVEMSLDCNDLILVDPIGATGTCGEYSCPH
          QFTSPQENKQGFANWAMFVLEQHLKCKKADPENSYTATLHAKGTKKIAQKVGEE
          GVETLAAVAQDAKAEVISELDLVYHLTVLLHNDQLQWYEIIAKLQHRHQIGLHPG
          GNK"
gene      5323..5475
CDS       /gene="HI0476"
          5323..5475
          /gene="HI0476"
          /note="hypothetical protein; identified by Genemark;
          putative"
          /codon_start=1
          /transl_table=11
          /product="H. influenzae predicted coding region HI0476"
          /protein_id="AAC22144.1"
          /db_xref="GI:1573454"
          /translation="WVVKCIDKQOQNLGNILFLLKQYSEKDSKKFTIYKFLQTVN
          YTIQLS"
gene      5505..6707
CDS       /gene="HI0477"
          5505..6707
          /gene="HI0477"
          /note="similar to GB:M23240 SP:P18199 PID:148089 GB:U00096
          PID:1736569 percent identity: 47.24; identified by
          sequence similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="tyrosine-specific transport protein (tyrP)"
          /protein_id="AAC22135.1"
          /db_xref="GI:1573455"
          /translation="NKNKTVGSLLVAGTWIGMGLMPLTSAGIGFQFTLVLLGLWA
          LUTFSALLVEVYQTAESDAGTLAGQYFGTKRIATAVLIIIFLVALTAAYISGGG
          SLKDLPSFGDKSVLLETVIFGSFVIGTHSVKINRVLFVFMVAFAVLSML
          PEIKFDNLMATPIDKALIIISAPVFTAFGPHGSPSLNKYLDGNVKALRESILVGA
          ITLCAYILWQLSTHGLLTQNEFLQILKEDATLNGLVKATFAITGNSVIVASVAKLFSLI
          "
Alignment Scores:
Pred. No.: 1,le-87      Length: 13828
Score:      1173.00      Matches: 221
Percent Similarity: 91.00%      Conservative: 42
Best Local Similarity: 76.47%      Mismatches: 26
Query Match: 82.09%      Indels: 0
DB:      1      Gaps: 0

US-09-545-199F-4 (1-289) x U32730 (1-13828)

QY      1 MetAlaGlyAlaLysGluLeuArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
Db      9610 ATGGCAGGTGCAAAAGAGATAAAACCAAAATTGCCAGTGATACACAAATAATC 9551
QY      21 ThrLysAlaMetGluMetValAlaLysSerMetArgLysThrGlnGluArgMetSer 40
Db      9550 ACTAAGGCAATGGAATGTGGCAACCTCGAAATTCGTAACCAACCGAGGATCGATGGCT 9491
QY      41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db      9490 GCATCTCGTCGCTATTCTGAAACTATCGTACGTTATTACTCATGTGCTAGGCAAGT 9431
QY      61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80

```

```

Db      9430 ATCGGTTATAACATCGTTCTTAGTGAGCGCGAAGTCAAGAAAATCGGTATCTGGTT 9371
QY      81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db      9370 ATTTCACACAGATCGTGGGATGTGTGGTGGTTAAATGTTAAATTTATTTCAAAACACACTT 9311
QY      101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db      9310 AACCAATAAAAAATTCGAAAGAACAAATAATTTCTACAGATTTGGCTTTAATAGTTCA 9251
QY      121 LysSerIleAsnPhePheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db      9250 AAGGGGATAGTTTTTCGCTTTCCTTTGGATTATATCAAAAGTCAGCTTCTCGTTTA 9191
QY      141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db      9190 GCGGATACGCCCTCTAGAGAAATTAATGTGTGGCAATACAAATGTTTGTATGCTTAT 9131
QY      161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db      9130 CGTAATGGTGAATTCATGCAGTTTATATTCATACATAAATTTGTTAATACGATGTCG 9071
QY      181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db      9070 CAAAACCTTGTTGACACAAATAGTTCTTTTACAGAAATCTAAAGACGATCATTTAAAT 9011
QY      201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db      9010 GAAAGCAACACAGACTTGGGATATCTTTATGAGCCAGAACCAAAAGTAGTACTATTAGATAGC 8951
QY      221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240
Db      8950 CTTTATAGTTTCGTTATTAGTCCCAAAATTTATCAAGCGGTTGTAGATAATGTAGCTTCA 8891
QY      241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db      8890 GAACAGCGCTCGAATGATGATGAACAGCAACCGGATATGTCAGGTAATTTAATTT 8831
QY      261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db      8830 AATGATCTCGGTTGGTGATACAAACAACTCTCTCAAGCAAGTATCACAAATGAATTTGAAT 8771
QY      281 GluIleValAlaGlyAlaAlaAlaIle 289
Db      8770 GAATCGTAGCGGTCGCGCGCGGATT 8744

```

RESULT 6  
AR274513\_04/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS AR274513 Accession AR274513  
Fragment Name Begin End

```

AR274513_00 1 110000
AR274513_01 100001 210000
AR274513_02 200001 310000
AR274513_03 300001 410000
AR274513_04 400001 510000
AR274513_05 500001 610000
AR274513_06 600001 710000
AR274513_07 700001 810000
AR274513_08 800001 910000
AR274513_09 900001 1010000
AR274513_10 1000001 1110000
AR274513_11 1100001 1210000
AR274513_12 1200001 1310000
AR274513_13 1300001 1410000
AR274513_14 1400001 1510000
AR274513_15 1500001 1610000
AR274513_16 1600001 1710000
AR274513_17 1700001 1810000
AR274513_18 1800001 1830121

```

Continuation (5 of 19) of AR274513 from base 400001 (AR274513 Sequence 1 from patent US 6  
Alignment Scores:

Pred. No.: 1,42e-86 Length: 110000  
Score: 1173.00 Matches: 221  
Percent Similarity: 91.00% Conservative: 42  
Best Local Similarity: 76.47% Mismatches: 26  
Query Match: 82.09% Indels: 0  
DB: 6 Gaps: 0

US-09-545-199f-4 (1-289) x AR274513\_04 (1-110000)

QY 1 MetAlaGlyAlaLysGluLeuThrLysLeuValSerValLysSerThrGlnLysLeu 20  
DB 101934 ATGCGAGTGCAGAAAGAGATAAAACCAAAATTCGAGTACAGATACACAAAATC 101875

QY 21 ThrLysAlaMetGluMetValAlaLysSerValLysSerThrGlnLysLeu 40  
DB 101874 ACTAAGCAATGAAATGCGCAACCTCGAAATGCGTAAACGCGAGTATCGT 101815

QY 41 SerSerArgProTyrSerGluThrLysLeuValSerValLysSerThrGlnLysLeu 60  
DB 101814 GCATCTCGTCGATTCGAAACTATCGTAACTATTCGATGATGCTTAAGGCAAGT 101755

QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
DB 101754 ATCGGTATTAACATCCGTTCTTAGTGCGGAGTGAAGAAATCCGTTATCTGTT 101695

QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuValLysLeuPheLysThrValLeu 100  
DB 101694 ATTTCAACAGATCGTGGGATGTCGTTGATTAATGTTAATTTATTTCAAAACACACTT 101635

QY 101 AsnGluMetLysGluTyrLysLysAspValSerValLysSerThrGlnLysLeu 120  
DB 101634 AACCAATAAAATTTGGAAGAACAAATTTCTACAGATTTGGCTTAATAGTTCA 101575

QY 121 LysSerIleAsnPheGlnSerLeuGlyLysLeuValLysLeuThrGlnAspSerGlyLeu 140  
DB 101574 AAAGGATAGTTTTCGTTCCGTTTGGATTAATATCAAAAGTCAGCTTCTGTTTA 101515

QY 141 GlyAspThrProSerValGluGlnLeuLysLeuValLysSerMetIleAspAlaTyr 160  
DB 101514 GCGGATACGCGCTCTAGAACAAATTTGTTGTCGCAATACAAATGTTGATCTAT 101455

QY 161 LysLysGlyGluValAspValLysLeuValTyrAsnLysPheLysLeuThrMetSer 180  
DB 101454 CGTAATGGGAAATGATGCAAGTTTATATTCGATACAAATTAATTTGTTAAGCATGTCG 101395

QY 181 GlnLysProValLysLysLeuLeuProLeuProGluLeuAspAsnGluLeuGly 200  
DB 101394 CAAAGCGTGTGTACAAATAGTTCTTATCCAGAACTCAAGAACTCAAGCAATTAAT 101335

QY 201 GluArgLysGlnValTyrAspTyrLysLeuProAspAlaLysValLysLeuAspAsn 220  
DB 101334 GAAAGACACACAGACTGGGATTCCTTATGAGCAGAACCAAAAGTACTATTAGATAGC 101275

QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuLysLeu 240  
DB 101274 CTTTATGTTCTGTTATTTAGATGTCGAAATTTATCAAGCGTTGATGATGATTTCA 101215

QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaLysLeuThrAspAsnAlaGlyAsnLeu 260  
DB 101214 GAAAGCGGCTCGAATGATGAGCAATGAAGCAGCAACGCAATGATGATGATTTAAT 101155

QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
DB 101154 AATGATCTGCGGTTGGTGTACAAAGAGCTGTCAGCAAGTATCAAGATGAATGAAT 101095

QY 281 GluLeuValAlaGlyAlaAlaLeu 289  
DB 101094 GAAATCGTAGCGGTCGCGCGCGATT 101068

RESULT 7  
AR274513\_05/c  
WFCOMMENT

Sequence split into 19 fragments LOCUS AR274513 Accession AR274513

Fragment Name Begin End  
AR274513\_00 1 110000  
AR274513\_01 100001 210000  
AR274513\_02 200001 310000  
AR274513\_03 300001 410000  
AR274513\_04 400001 510000  
AR274513\_05 500001 610000  
AR274513\_06 600001 710000  
AR274513\_07 700001 810000  
AR274513\_08 800001 910000  
AR274513\_09 900001 1010000  
AR274513\_10 1000001 1110000  
AR274513\_11 1100001 1210000  
AR274513\_12 1200001 1310000  
AR274513\_13 1300001 1410000  
AR274513\_14 1400001 1510000  
AR274513\_15 1500001 1610000  
AR274513\_16 1600001 1710000  
AR274513\_17 1700001 1810000  
AR274513\_18 1800001 1910000  
AR274513\_19 1900001 2010000

Continuation (6 of 19) of AR274513 from base 500001 (AR274513 Sequence 1 from patent US

Alignment Scores:  
Pred. No.: 1,42e-86 Length: 110000  
Score: 1173.00 Matches: 221  
Percent Similarity: 91.00% Conservative: 42  
Best Local Similarity: 76.47% Mismatches: 26  
Query Match: 82.09% Indels: 0  
DB: 6 Gaps: 0

US-09-545-199f-4 (1-289) x AR274513\_05 (1-110000)

QY 1 MetAlaGlyAlaLysGluLeuThrLysLeuValSerValLysSerThrGlnLysLeu 20  
DB 1934 ATGCGAGTGCAGAAAGAGATAAAACCAAAATTCGAGTACAGATACACAAAATC 1875

QY 21 ThrLysAlaMetGluMetValAlaLysSerValLysSerThrGlnLysLeu 40  
DB 1874 ACTAAGCAATGAAATGCGCAACCTCGAAATGCGTAAACGCGAGTATCGTATGTT 1815

QY 41 SerSerArgProTyrSerGluThrLysLeuValSerValLysSerThrGlnLysLeu 60  
DB 1814 GCATCTCGTCGATTCGAAACTATCCGTAACCTATTTAGTATGCTGCTTAAGCAAGT 1755

QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
DB 1754 ATCGGTATTAACATCCGTTCTTAGTGCGGAGTGAAGAAATCCGTTATCTGTT 1695

QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuValLysValLysLeuPheLysThrValLeu 100  
DB 1694 ATTTCAACAGATCGTGGGATGTCGTTGTTAATGTTAATTTATTTCAAAACACACTT 1635

QY 101 AsnGluMetLysGluTyrLysLysAspValSerValLysSerThrGlnLysLeu 120  
DB 1634 AACCAATAAAATTTGGAAGAACAAATTTCTACAGATTTGGCTTAATAGTTCA 1575

QY 121 LysSerIleAsnPheGlnSerLeuGlyLysLeuValLysLeuThrGlnAspSerGlyLeu 140  
DB 1574 AAAGGATAGTTTTCGTTCCGTTTGGATTAATATCAAAAGTCAGCTTCTGTTTA 1515

QY 141 GlyAspThrProSerValGluGlnLeuLysLeuValLysSerMetIleAspAlaTyr 160  
DB 1514 GCGGATACGCGCTCTAGAACAAATTTGTTGTCGCAATACAAATGTTGATCTAT 1455

QY 161 LysLysGlyGluValAspValTyrAsnLysPheLysLeuValLysSerMetSer 180  
DB 1454 CGTAATGGGAAATGATGCAAGTTTATTCGATACAAATTAATTTGTTAAGCATGTCG 1395

QY 181 GlnLysProValLysLysLeuLeuProLeuProGluLeuAspAsnGluLeuGly 200  
DB 1394 CAAAGCGTGTGTACAAAGAGCTTCTTACCAGAAATCTAAAGACGATCAATTAAT 1335

QY 201 GluArgLysGlnValTyrAspTyrLysLeuProAspAlaLysValLysLeuAspAsn 220

```

1334 GAAAGCAACAGACTGGGATTATCTTTATGATGCCAGAACCAAAAGTACTATTAGATAGC 1275
221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240
1274 CTTTATGTTGGTTATTAGTCCCAATTTATCAAGCGGTTGATGATGATGATGATGATCA 1215
241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLe 260
1214 GAACAGCGCGTCAATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1155
261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
1154 AATGATCTGCGTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
281 GluLeuValAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 289
1094 GAAATCGTAGCGGTCGCGCGCGGAT 1068

RESULT 8
BD252190
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BD252190.1 GI:33061960
JP 2002541790-A/74.
Actinobacillus pleuropneumoniae
Actinobacillus pleuropneumoniae
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Actinobacillus.
1 (bases 1 to 867)
Lowery, D.E., Fuller, T.E. and Kennedy, M.J.
Anti-Bacterial vaccine compositions.
Patent: JP 2002541790-A 74 10-DEC-2002;
PHARMACIA AND UPJOHN CO
OS Actinobacillus pleuropneumoniae
PN JP 2002541790-A/74
PD 10-DEC-2002
PF 06-APR-2000 JP 2000611649
PR 09-APR-1999 US 60/128689,10-SEP-1999 US 60/153453 PI
DAVID E LOWERY, TROY E FULLER, MICHAEL J KENNEDY PC
C12N15/09, A61K39/02, A61K39/102, A61K39/395, A61P31/04, C07K14/285, PC
C07K16/12
PC C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N15/01, C12P21/02 PC
, C12P21/08, C12Q1/18,
PC GOIN33/15, GOIN33/50, GOIN33/53, GOIN33/569, GOIN33/577, C12N15/00,
PC C12N15/00,
PC C12N5/00
CC atpG
FH Key
FT CDS Location/Qualifiers
i .867
/organism="Actinobacillus pleuropneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:715"

ORIGIN
Alignment Scores:
Pred. No.: 8-68e-86 Length: 867
Score: 1132.50 Matches: 215
Percent Similarity: 89.97% Conservative: 45
Best Local Similarity: 74.39% Mismatches: 28
Query Match: 79.25% Indels: 1
DB: 6 Gaps: 1

US-09-545-199F-4 (1-289) x BD252190 (1-867)

QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
DB 1 ATGGCAGGTGGCAAGAGATAGAACCAAAATTCGAAGTGTGAATAATCTCAAAAATC 60
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40

```

```

Db 61 ACCAAAGCAATGGAAATGGTTGCTACCTCTAAATATCGTAAATGCGTAAAGCGAGCGATGGCT 120
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db 121 GCCAGTCGCTCTATTTCGGAACCAATCCGTAAGTGATAGCCATATTTCGGAAGGAAGC 180
QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Db 181 ATTGGTTATAAGACCCGCTTTTAACTGAACCGTGATATTAAAAAGTAGGCTATCTTGTTC 240
QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db 241 GTTTCGACCGATCGCGGTTTATCGGTTGGCTTTAATATCAATTTATTCAAGCGATTTC 300
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db 301 AATGAATTTAAACGCTGGAAGATAAAGACGTTAGTGTGAGCTTGCTTTAGTAGGTCG 360
QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db 361 AAAGCGTAGCTTTTACCAAAATCTAGGCTTAAACGTGAGATCTCAAGTAAACGGGATTA 420
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 421 GCGGATAATCCGAAATGGAACGATATCGTGGCGCAGTTAATCAATGATTAATCGGTTTC 480
QY 161 LysLysGlyGlnValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 481 CGAACGGAGAGTGGATCGGTTTACGTCGCTTAAACCGCTTTTGAATAACGATGTCA 540
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 541 CAAAAACCTGTTATCGCACAGTTACTTCGTTACCTAACTAGATGACGATGAATTA --- 597
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 598 GATACGAAAGGTTCTAGGATTTATTTATGAACCGAATCCCAAGTTTATTGGATAGT 657
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240
Db 658 TTACTTGTTCGTTATTAGAACTCAGGTATACAGCGGTTGATAGATACCTAGCTTCT 717
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db 718 GAACAGCGCGTCGAATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db 778 GATGAATTACAAATTAGTGTATACCAAGCTCCCAAGCAAGCATTACAAATGAATTAAC 837
QY 281 GluIleValAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 289
Db 838 GAAATGTTGCGGTCGCGCAGCAATT 864

RESULT 9
AX553798
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .867
AX553798
Sequence 132 from Patent WO02075507.
AX553798.1 GI:25897796
Actinobacillus pleuropneumoniae
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Actinobacillus.
Lowery, D.E., Fuller, T.E. and Kennedy, M.J.
Anti-Bacterial vaccine compositions
Patent: WO 02075507-A 132 26-SEP-2002;
Pharmacia & Upjohn Company (US)
Location/Qualifiers
1. .867

```

/organism="Actinobacillus pleuropneumoniae"

/mol\_type="unassigned DNA"

/db\_xref="taxon:715"

gene 1..867

CDS 1..867

/gene="atpG"

/gene="atpG"

/codon\_start=1

/transl\_table=11

/protein\_id="CAD58183.1"

/db\_xref="GI:2589779"

YSETIRKVISIAKSGYKHPFLTERDKKGVLYVSTDRGLCGGLNLPLKATLINE  
FKTKWDKVSVELGLVSGKVSFYONLGNVRSQVTGLGNPWERIVCAVNMENINAF  
RNGEVDVAVYNNFNTMSQKPVIAOLLPLKLDDELDTKGSWDYITTEPNPOVLLD  
SLLYLETOVYQAVVDNLASEQARVAVKAAIDNAGLILDELQLVYNKARQASITN  
ELNEIVAGAAAI"

# ORIGIN

## Alignment Scores:

Pred. No.: 9.68e-86 Length: 867  
Score: 1132.50 Matches: 215  
Percent Similarity: 89.97% Conservative: 45  
Best Local Similarity: 74.39% Mismatches: 28  
Query Match: 79.25% Indels: 1  
DB: 6 Gaps: 1

US-09-545-199f-4 (1-289) x AX553798 (1-867)

QY 1 MetAlaGlyAlaLysGluLeuAArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
DB 1 ATGCAGGTCGGAAGAGATAGAACCAAAATTCGAAGTGTGAATAAATCAAAAATC 60  
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnLysArgMetSer 40  
DB 61 ACCAAGCAATGGAATGTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120  
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
DB 121 GCCAGTCGTCCTTATTCGGAAACAAATCCGTAGGTGATTAGCCATATTGCCAAGGAAGC 180  
QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
DB 181 ATTGGTTATTAAGCACCGCTTTTAACTGAACGTGATATTAAAGATAGGTCTCTCTCTC 240  
QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
DB 241 GTTTCACCGATCCCGGTTATTCGGTGGCCCTTAATCAATTTATTCAAAAGCACTTG 300  
QY 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLysIleGlySer 120  
DB 301 AATGAATTTAAACGTCGGAACGATAAAGACGTTAGTGTGAGTTGCTTTAGTAGGCTCG 360  
QY 121 LysSerIleAsnPheGlnSerLeuGlyLysIleLysIleLeuThrGlnAspSerGlyLe 140  
DB 361 AAAGCGTAAAGCTTTTACCAATCTAGCCTTAAACGTGAGATCTCAAGTAAACGGGATTA 420  
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160  
DB 421 GCGGATAATCCGGAATGGAACGATCTCGTGGCGGCGAGTTAATGAATGATTAATGCGTTC 480  
QY 161 LysLysGlyGluValAspValValLysValLysValLysLysPheIleAsnThrMetSer 180  
DB 481 CGAAACGGAGAGTGGATCGCGTTTACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540  
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
DB 541 CAAAACCTGTTATCCACAGCTTACTCCGTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 597  
QY 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
DB 598 CATACGAAGGTTTCATGGGATTTATTTATGAACCGAATCCCAAGTTTATTTGGATAGT 657

QY 221 LeuLeuValAArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240  
DB 658 TTACTTGTTCGTTATTATTAGAACTCAGGTATAGCAAGCAGTTGTAGTAACCTAGCTTCT 717  
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260  
DB 718 GAACAAGCGGCTCGAATGTTAGCGATGAAGCGGCAACGATTAATCGGGTACATTAATC 777  
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
DB 778 GATGAATTAACAATTAGTGTATTAACAAGCTCGCAAGCAAGCATTACAATGAATTAAC 837  
QY 281 GluLeuValAlaGlyAlaAlaAlaIle 289  
DB 838 GAAATTTGTCGGGTGCGCGACCAATT 864

# RESULT 10

AE017151

LOCUS

DEFINITION

Haemophilus ducreyi strain 35000HP section 1 of 6 of the complete

Genome.

ACCESSION

AE017151 AE017143

VERSION

AE017151.1 GI:33147502

KEYWORDS

SOURCE

ORGANISM

Haemophilus ducreyi 35000HP

Haemophilus ducreyi 35000HP

Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

Pasteurellaceae; Haemophilus.

1 (bases 1 to 304558)

Munson,R.S. Jr., Ray,W.C., Mahairas,G., Sabo,P., Mungur,R.,

Johnson,L., Nguyen,D., Wang,J., Forst,C. and Hood,L.

The Complete Genome Sequence of Haemophilus ducreyi

Unpublished

2 (bases 1 to 304558)

Munson,R.S. Jr., Ray,W.C., Mahairas,G., Sabo,P., Mungur,R.,

Johnson,L., Nguyen,D., Wang,J., Forst,C. and Hood,L.

Direct Submission

Submitted (04-JUN-2003) Pediatrics, Columbus Children's Research

Institute and The Ohio State University, 700 Children's Drive,

Columbus, OH 43205, USA

LOCATION/Qualifiers

1..304558

/organism="Haemophilus ducreyi 35000HP"

/mol\_type="genomic DNA"

/strain="35000HP"

/db\_xref="taxon:233412"

296..2188

/gene="gidA"

/locus\_tag="HD0001"

/codon\_start=1

/transl\_table=11

/product="glucose inhibited division protein A"

/protein\_id="AAP95025.1"

/db\_xref="GI:33147503"

/translation="MIYPEHYDVIYVGGHAGTEAALAPARMGLKTLILTHNVDITLQ

MSNPAIGGIGKGLVKEIDAMGLMAIATDHAGIQFRLNSKSGPAVRATRAQADRV

LYQSVRTALENOQNLAIQQEVDVILLVNHASGVTKGLTFKAKTILTAGTFLA

GKTHIGLDNVAGGRADGPAATRLAERLDNLKVDLKTGTPPRDLARTINFEVLAKQ

PGSEVLPVMSPGMEGQHPQIPCYITHTNHQTHDIIRANLDRSPMTGIIIEGIPRY

CPSEIDKWMFGDRNGHQIYLEPGTLTIEVYPNGISTSLPFDVQMAVNSMKGLENT

RIIQAVALYEYDFDPRDLKPTLETKAIEGLFFAGQINGTTGVEAAAGGLAGLNAA

LYQAKPSFFPTDVATGVLDVLCITLGTKEPVFTSRAEYRLLLEDNADRLTLP

IAHQGLIDEARWQFNQENIERERELKQIWIHPQSAVYDVAVNQLVNSPLREVN

GEDLLRRPINAKLTEIAAFAPALDDQRAEOVEIAIKYQGYIEHQHEIERKHRE

NLIIPAEFYDYNKVDLSNEVRKLMQHRPISIGQASRISGITPAAISILLVNLKQGM

LKRGEL"

2293..2904

/gene="gidB"

/locus\_tag="HD0002"

2293..2904

gene

CDS

```

/gene="gidB"
/locus_tag="HD0002"
/locus_start=1
/codon_start=1
/transl_table=11
/product="glucose inhibited division protein B"
/protein_id="AAP95026.1"
/db_xref="GI:33147504"
/translation="MQKDLRLVQADIGLITRQKEQLGFGVHLLHKNNKAYNLTSVR
DPEMLVKHIMDSLVSVHLOGSFIDVGTGFGGLPIAIANPKQVLLDSLGKRI
TFIKQVRLKELAINTVPLSRVEQYTAQFDGVLGRFASLDMLNWCYALPNSGKF
YALKGVDTGELQAVQPKLVDAIKLNPQLVGRHLIVLAK"
3220..3603
/locus_tag="HD0003"
3220..3603
/locus_tag="HD0003"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAP95027.1"
/db_xref="GI:33147505"
/translation="MSAVINKAKOHLMAKLGSLFAICMLLILGSLSPSLGSG
CLASFPYCLFVYVWIEFKSAKQSKMAFYRGEGLKWLATILLVVAFAKLIPLHRV
LPFVGFEVALLNNVPPVQKRTN"
3629..4417
/locus_tag="HD0004"
3629..4417
/locus_tag="HD0004"
/codon_start=1
/transl_table=11
/product="ATP synthase A chain"
/protein_id="AAP95028.1"
/db_xref="GI:33147506"
/translation="MAVNTAEYGHGLSGLSGDFWAVHLDLPFSLVAGVLFVVF
SRVAKNATGVGKGLQVEMVWVWGLVKDNFGPHMTAPLALIFCWVFMNAL
DLPVDFPLQLANMFGIHLRAVTPADISATLGSICVFGILLPTVKSQGFNGLAKE
VILHPNHWAPTPVNFILETVLLAKPISLAFRLFGNMYAGELIFILIAVMYADNPA
LQALGIPLWVAIFHILVITLQAFIFWMLTIVLSIAYNKADH"
4538..4792
/locus_tag="HD0005"
4538..4792
/locus_tag="HD0005"
/codon_start=1
/transl_table=11
/product="ATP synthase C chain"
/protein_id="AAP95029.1"
/db_xref="GI:33147507"
/translation="MESVITATIGASILLAPALGTAIGFALIGKGFLESSRQPEL
ASSLQTKFTVINGLLDAIAMVAGISLLFIFANFIDLLK"
4846..5316
/locus_tag="HD0006"
4846..5316
/locus_tag="HD0006"
/codon_start=1
/transl_table=11
/product="ATP synthase B chain"
/protein_id="AAP95030.1"
/db_xref="GI:33147508"
/translation="MNLNATLIGOLIAFALPVAFCKYVWPEPLIKAIERSQANIAL
SSAKKAEQDSKVLQAEQETIKAKAEQKIDLAKRNEILLESVQDAEVEQRRII
EQGHAEIDSEKRVQEELRQKVAALAVAGAEIRVGRSVDTAANNIDILKLVAEI"
5330..5863
/locus_tag="HD0007"
5330..5863
/locus_tag="HD0007"
/codon_start=1

```

```

/transl_table=11
/product="ATP synthase delta chain"
/protein_id="AAP95031.1"
/db_xref="GI:33147509"
/translation="MSELSTVARPYAKAADFALFEGQLDKWQEMLOFSALVAENEQV
YETITSSALSGQISFIQICDDQDQYGNFIRVMAENKELAILPEVFNFRNALRAE
YKAIKKVIVSATLSLSQAEEKIAKAMEKRLQKQVRLTTEIDSTLLTGIIIKYDDIVI
DSSSRGQNLRLTQALS"
5876..7417
/locus_tag="HD0008"
5876..7417
/locus_tag="HD0008"
/codon_start=1
/transl_table=11
/product="ATP synthase alpha chain"
/protein_id="AAP95032.1"
/db_xref="GI:33147510"
/translation="MQLNSTEISELIIKRIAQFNWVSEAOSTGTIVSVSDGIIRHGL
ADVQGEIMELFNGRIYALNALNRDSVAGVVMGYADLAEGMTVKCTGRILEVPVGKG
LLGRVNTLGPIDGKEIDNDGSPVEVIAPQVDSVDQPVQGTGKAVDSMPVIG
RQRELLIGDRQITGKTSILVIDAIQAORDSGIKCIYVALGOKASTIANVVKLSHGAL
ANTIWWASASEAALQVLAIPYSGCAMEXERDEGADALIVDDLKQAVAYQISLL
LRPPGREAFGQDVFLYHSLRLLERAARVNAEYVERFTNGAVTGRGSLTALPIETQA
GDVSAFVPTNVISITDQIFLESIFNSGIRPAVNPGISVSRVSGSAQTQVVKLSGG
IRTALAQRLELAFAFASDLDLDRKQLSHGQKVTLLKQKQFAPMPVSEQALVLPFA
AFGYLDGVLELIRISSFESALLAYANSNHADFMKDLAKSGNFNDISIQOELKTIVENFK
KNGAW"
7441..8307
/locus_tag="HD0009"
7441..8307
/locus_tag="HD0009"
/codon_start=1
/transl_table=11
/product="ATP synthase gamma chain"
/protein_id="AAP95033.1"
/db_xref="GI:33147511"
/translation="MAGAKSIRTKIASVRNTOKITKAMEMVATSKMRKTQERMAGRP
YSETIRKVIASHIAGSIGYKHPFLIERDVKVGLVISTDRGLCGGINLFPKTLNE
FKAWKDXSVELGVLGSGVSGVYIGLKVRAHITGLGDSPEMERIVGAVNEMINAY
RNGEVDVVCYAVNRFENTMSQKTVIAQLPLPLENDELETKCSWDYLYEPNPQVLLD
SLIRYLETOVYQAVVDNLASEQAARVMAKATDNAGALIDELQLVYNKARQASITN
ELNEIVAGAAI"
8323..9696
/locus_tag="HD0010"
8323..9696
/locus_tag="HD0010"
/codon_start=1
/transl_table=11
/product="ATP synthase beta chain"
/protein_id="AAP95034.1"
/db_xref="GI:33147512"
/translation="MATGKIVQIIGAVIDVERPQDAPKVYDALKVESGLTLEVOQL
GGGLVRCIALGTSGLKRGKLVKNTGNPLQVPGTKLGRIMNVLGEPIDEKGPGEZ
ARWIDIHRAAPSVEEQSNTELETGTIKVIDLICPFKAGGKGLFGAGVGKVTNNMEL
IRNIAIEHSGYSYFAGVAGTEREGNDYHEMIDSNVLDKSVLYVGOMNEPPGNLRVA
LTGLTWAKFRDEGRDVLFFVNYRITLAGTEVSALLGRMPSAVGYOPTLAEEGVL
QERITSTKTSITSVQAVYPADLTPSPATTFALDSTVVSLSNIASLGTIYPAVDP
LDSTSLQPLPLVVGEEHYNVARVGQTLQRYKELDIIAILGMDSELDLKLVSRRAR
KIERFLSQPFFVAEVTGSPGVSLKDTIRGFGILSGEFDHIFEQAFYVAGSIDEV
VERASNM"
9736..10155
/locus_tag="HD0011"
9736..10155
/locus_tag="HD0011"

```

## Alignment Scores:

Pred. No.: 1.02e-81 Length: 304558  
 Score: 1121.50 Matches: 212  
 Percent Similarity: 89.97% Conservative: 48  
 Best Local Similarity: 73.36% Mismatches: 28  
 Query Match: 78.48% Indels: 1  
 DB: 1 Gaps: 1

US-09-545-199F-4 (1-289) x AE017151 (1-304558)

QY	1	MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle	20
DB	7441	ATGCGAGGTGCTAAAGAGATTAGAACCAAAATTCGAAGTTCGTAATACGCAAAATTT	7500
QY	21	ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer	40
DB	7501	ACGAACCAATGMAATGGTGGTACCTCTAAATTCGTAAATCGTAAACGCAAGAGGTATGCA	7560
QY	41	SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr	60
DB	7561	GCAGGTGCTCCATATTACAGACGATCCGTAAAGTAATTAGCCATATCGCAAAAGGAAGC	7620
QY	61	IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal	80
DB	7621	ATTGGTTATAGCACCCGCTTTTATAGCGGTGATGTTAAAGATGTTAAAGATGTTAGTT	7680
QY	81	ValSerThrAspArgGlyLeuGlyGlyLeuAsnValAsnLeuPheLysThrValLeu	100
DB	7681	ATCTCTACCGACCGCGTTTATGCGCGTCTTATATCAATTTATTCAGACACCTTG	7740
QY	101	AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer	120
DB	7741	AATGAATTTAAGCGTGAAGAGATAAGACGTCAGTGTGAATTTGGCTTAGTGGCTCA	7800
QY	121	LysSerIleAsnPhePheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle	140
DB	7801	AAGGAGTAGTACCTTTTACCAATCTATGCTTAAAGTTAGGCTCATATTACAGTTTA	7860
QY	141	GlyAspThrProSerValGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr	160
DB	7861	GGTGATTCACCTGAAATGAAAGCGTATTGTTGGTCAGTTAATGAATGAATGATCTAT	7920
QY	161	LysLysGlyGluValAspValValTyrLysValTyrAsnLysPheIleAsnThrMetSer	180
DB	7921	CGCAATGGTGAAGTGATGATGTTTGTATGCTTATATCGTTTGAATATACATGCTCA	7980
QY	181	GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnGluLeuGly	200
DB	7981	CAAAACACGTTATTGCAATATCTACCGTTACCTAAATTAGAAATGATGAATTA--	8037
QY	201	GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn	220
DB	8038	GAACAAATGTTCTTGGATTTATCTTACAGCGCTAACCCACAGTATTATTGGATAGT	8097
QY	221	LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer	240
DB	8098	TTACTATTTCGTTATTAGAACACAGGTATATCAAGCAGTTGTCGATAACCTTGCTTCT	8157
QY	241	GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle	260
DB	8158	GAACAGCGCTCGAATGTTAGCAATGAGCGCCACACATATTCGGGAGCATTAATT	8217
QY	261	AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn	280
DB	8218	GATGAGTTACAAATTAGTGTATPACAAAGCTCGTCAAGCAAGTATTACAAATGAATTAAT	8277
QY	281	GluIleValAlaGlyAlaAlaIle	289
DB	8278	GAAATTGTCAGGTGCTGCTCAATT	8304

RESULT 11

AX553832

LOCUS

866 bp

DNA

linear

PAT 27-NOV-2002

## DEFINITION

Sequence 166 from Patent WO02075507.

## ACCESSION

AX553832 GI:25897830

## VERSION

AX553832.1 GI:25897830

## KEYWORDS

Mannheimia haemolytica

## SOURCE

Mannheimia haemolytica

## ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

## REFERENCE

1 Lowery, D.E., Fuller, T.E. and Kennedy, M.J.

## AUTHORS

Anti-bacterial vaccine compositions

## TITLE

Patent: WO 02075507-A 166 26-SEP-2002;

## JOURNAL

Pharmacia &amp; Upjohn Company (US)

## FEATURES

Location/Qualifiers

1..866 /organism="Mannheimia haemolytica"

/mol\_type="unassigned DNA"

1..864 /db\_xref="taxon:75985"

1..864 /gene="atpG"

/codon\_start=1

/protein\_id="CAD58199.1"

/db\_xref="GI:25897831"

/translation="MAGAEIRTKIASVRNTOKITKAMEMVAASKKRKTQERMAASRP

YAESRKAISHIAKNIIEYKHPFLTPRPVKVGLVYSTDRGLCGGGINILKTVLHE

LKEKQDQGVKSLAVGVNGKISFPNMGLEIKGHINGLGTDPAMEDLVGVNGVNRVY

REGEDIVVYVNRFINTMKSQKPTVQQLPLPALENDSELTQTSWDLYLPEPFCALHD

SILVRYLESQVYQAVDNLASEQARVAVKATDAGNLINEQLQVINKARQASITN

ELNEIVAGAAI"

## ORIGIN

## Alignment Scores:

Pred. No.: 3.44e-81 Length: 866  
 Score: 1077.50 Matches: 208  
 Percent Similarity: 87.89% Conservative: 46  
 Best Local Similarity: 71.97% Mismatches: 34  
 Query Match: 75.40% Indels: 1  
 DB: 6 Gaps: 1

US-09-545-199F-4 (1-289) x AX553832 (1-866)

QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20

DB 1 ATGCGAGGTGCTAAAGAGATAAGAACCAAAATTCGAAGTTCGTAATACACAAAAATTT 60

QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40

DB 61 ACCAAACGCGATGGAATGTTGCCGATCAAAATCGCTAAACCCACGAGGTATGGCG 120

QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60

DB 121 GCTTCTCGCCCTTATGCTGAAAGTATTCGACGCAATTAGCCATATTGCCAAAGTAAAC 180

QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal 80

DB 181 ATTGAGTATAACACCCATTTTGACCCACGCTCCGGTAAAAAAGTTGGCTATTAGTA 240

QY 81 ValSerThrAspArgGlyLeuGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100

DB 241 GTTTCACCATCCCGCTTTATGTGGTGGCTTAATATCAATTTATTATAACCGTTTAA 300

QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120

DB 301 CATGAATTTGAAGAAAGATGACCAAGGTGTTAAGTCTCGACTTGTGTGTCGGGAAT 360

QY 121 LysSerIleAsnPhePheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140

DB 361 AAGGGATTCTCTTTTAAACCAATGGGCTAGAGATTAAAGGTTCATATCAATGATG 420

QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160

```

Db 421 GGTGATACCGCAATGGAAGATTAGTCGGTATTGTTAATGATGTAAGTAATGCGCTAC 480
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 481 CGTGAAGCGCAATTCATGAGTGTATGTTGATATTAACCGTTTATATAACACGATGCA 540
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 541 CAAAAACCGACAGTACACAGTTGCTTCCCTTGCTGCACCTGGAAATGACTCATTA--- 597
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 598 GAGCAAACTGCTCTTGGGATTATCTCTATGAACCAATCCACAGCGTTATTAGACAGC 657
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAunLeuAlaSer 240
Db 658 TTACGTTGCTGTTATTAGAACTCAAGTTATCAGGAGTGGTAGATATCTTGGCTCT 717
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db 718 CAACAGGCTGCTCGAATGGTGGCAATGAACGACCAACCGATTAACGAGGTATCTGATT 777
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db 778 AATGAGTTACAGTTAGTGTATTAACAAAGCTCGTCAAGCAAGTATTACGAATGAATTAAT 837
QY 281 GluIleValAlaGlyAlaAlaAlaIle 289
Db 838 GAAATTGTGCGGCTGCGCGCAAT 864

RESULT 12
AR378271
LOCUS AR378271 876 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 3277 from patent US 6605709.
ACCESSION AR378271
VERSION AR378271.1 GI:40081453
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 876)
AUTHORS Breton,G.L.
TITLE Nucleic acid and amino acid sequences relating to Proteus mirabilis
for diagnostics and therapeutics
JOURNAL Patent: US 6605709-A 3277 12-AUG-2003;
FEATURES
Source
Location/Qualifiers
1..876
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.: 4,89e-75 Length: 876
Score: 1004.00 Matches: 194
Percent Similarity: 84.08% Conservative: 49
Best Local Similarity: 67.13% Mismatches: 44
Query Match: 70.26% Indels: 2
DB: Gaps: 1

US-09-545-199F-4 (1-289) x AR378271 (1-876)

QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
Db 13 ATGGCGCGGCGCAAGAGATAGTTCGAAGATCGCCAGTGTGCAAAACACACAGAGATC 72
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
Db 73 ACTAAGCGATGGAGATGCTCCCGCGTCGAAATGCGTAAACGCGAGGAACGATGGCA 132
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db 133 GCCAGCGCTCTTATGCGAAGAACCATGCGCAGTGTGATTGTCACCTTGGTAGGAAT 192

```

```

QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Db 193 CTGGAAATAAACAATCATACCTCGAAGCGTGAAGTTAARCGTGTGCGGTACCTGGTT 252
QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db 253 GTTTCGACCGACCGTGTGTTGTGTTGTTGAACATTAACTTGTTCAAAAAAGCTGCTG 312
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db 313 CGACACATGAAGAGTGTCTGATAAAGTGTGAAGTAGATTAGCCCTTGTGTTCT 372
QY 121 LysSerIleAsnPhePheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db 373 AAAGCCGTTTCATCTTTCCTCTGTAGGTGGGAATGTTGCGTCAAGTACAGGTATG 432
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 433 GGTGATGATCCGCAACTATCTGACCTTATCGGCCCAAGTCAATCATCATACCAATG 492
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 493 GACGAAGTGTGTTAGATAAAGTGTGTTGCGCAATGAAGTTATCATCATACCAATG 552
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 553 CAAGAGCTTAAATTTCTGCAAGTTTACCATTACCACAGGTGATGATGAAGAACTGAA 612
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 613 GAAAATCT-----TGGATTATCTGACGAACTGATCTTAAGACGTTACTTGATACT 666
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240
Db 667 TTACTCGCTCGTTATATTGAATCGCAGGTTTATCAGAGCGTCTGAGAGAACTTAGCGAGT 726
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db 727 GACGAGCGACGAGATGTCGCGATGAAGCGGTACAGATAACGGTGGTGAATCTGATT 786
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db 787 AAAGAACTGCAATTGGTGTACAAACACGACGCGCAAGCAAGCATTACTCAGGAACCTGACA 846
QY 281 GluIleValAlaGlyAlaAlaAlaIle 289
Db 847 GAAATTGTATCTGTCGACGCGCGTA 873

RESULT 13
BX571859 349483 bp DNA linear BCT 26-SEP-2003
LOCUS BX571859/c
DEFINITION Photorhabdus luminescens subsp. laumondii T101 complete genome;
segment 1/17.
ACCESSION BX571859 BX470251
VERSION BX571859.1 GI:36783455
KEYWORDS complete genome.
SOURCE Photorhabdus luminescens subsp. laumondii T101
ORGANISM Photorhabdus luminescens subsp. laumondii T101
REFERENCE 1
AUTHORS Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
Bocs,S., Boursaux-Eude,C., Chandler,M., Dassa,E., Derose,R.,
Derzelle,S., Freydisin,G., Gaudriault,S., Givaudan,A., Glaeser,P.,
Medigue,C., Lanois,A., Powell,K., Signier,P., Wingate,V.,
Zouine,M., Boemare,N., Danchin,A. and Kunst,F.
TITLE Complete genome sequence of the entomopathogenic bacterium
Photorhabdus luminescens
Nac. Biotechnol. 11 (1) (2003) In press
JOURNAL
REFERENCE 2
AUTHORS Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
TITLE Direct Submission

```



## JOURNAL

Submitted (23-APR-2003) L. Prangeul, Institut Pasteur, Genopole, 25  
rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:  
lfrangeul@pasteur.fr, fkuinst@pasteur.fr

## FEATURES

## source

1. 349483

/organism="Photobhabdus luminescens subsp. laumondii T101"  
/mol\_type="genomic DNA"

/strain="T101"

/db\_xref="taxon:243265"

223. 1622

/gene="dnaA"

/locus\_tag="plu0001"

223. 228

/gene="dnaA"

/locus\_tag="plu0001"

234. 1622

/gene="dnaA"

/locus\_tag="plu0001"

/codon\_start=1

/transl\_table=11

/product="Chromosomal replication initiator protein DnaA"

/protein\_id="CAE12296.1"

/db\_xref="GI:36783456"

/translation="WSLSLWQCCRLARQDELPAEFMSWIRPLQALSGNTLALYAPN  
RVLWVRKYNINGLNGLPCGAELPRLFEVGNKPVSONDSPQVTVTHPTAPPA  
PQTSVRPSMDTAVPELSVRSNPNKHTFDNFBGKSNQLAARAAQVADNPAGAY  
NPLFLYGGTGLKTHLHAGVNSIMARKANAKVVMHSERFQDMVKALQADNAIEBFK  
RYRSVDALLIDIQIOPFANKERSQEBEPHTNALLEGNOQIILTSRDVPEKINGVEDR  
LAKRSFOWGLTVAIEPELETVAIILMKKADENEIQLPGEVAFI AKELRSNVELEGA  
LNRVIANAFNGRAITIDFVEALRDLLALQELVTDNQKTVAEYIKIKVADLLSK  
RRRSVAPROMAMALAKELTNHSLPEIGDAPFGGRDHTTVLHACRKIEQUREESHDK  
EDFSNLITLS"

1627. 2727

/gene="dnaA"

/locus\_tag="plu0002"

1627. 2727

/gene="dnaA"

/locus\_tag="plu0002"

/codon\_start=1

/transl\_table=11

/product="DNA polymerase III, beta chain"

/protein\_id="CAE12297.1"

/db\_xref="GI:36783457"

/translation="WKFIIEREQLLKPQQVSSPLSGRPTLPILGNLLLVQTEGSLLL  
TGTDLEMMARVTLSPHEVGATVPARKEPFIWGLPDGAEISVELDGRLLVRS  
RSFSLTLPASDPNLDQWQSEVERSPQATLRLKRLIESTQFSEMAHQDVRVYLGWLF  
ETEGEELTAVDTHRLAVCSMGIGQLPSHSVIVPRKGVIELMLLLDGGDTPLQIQI  
GNNIRARVGFIFTSKLVDRFPDPRVLPKPNPKTLEANCMLKQAFSRAAILNE  
KFRGVLYFSENQLRI TANNPEQEAEEIIVDVSYQGAEMEIGFNVSYLDVLTLRCE  
EVRLLTDAISSVQIEDCANHNAAYVVMPL"

2749. 3840

/gene="recP"

/locus\_tag="plu0003"

2749. 3840

/gene="recP"

/locus\_tag="plu0003"

/codon\_start=1

/transl\_table=11

/product="DNA replication and repair protein"

/protein\_id="CAE12298.1"

/db\_xref="GI:36783458"

/translation="MTLRLLLRDFRNIAADLPATGFNPLVSGSGKTSVLEAIY  
TLGGRSFRIOAGRVILHGCDEFVLHGLGQOENERSIGLSKNRNGDSKVIDGS  
DGGKIALKMLPMQITPBGFTLLNGGPKYRAFIDMCGCFNRPFFSAWNLKRL  
KQNAALQVTRYSPQPMQDELIPLANQINQWGEYVNTIQDITNTCKOLFPEFTL  
SFSFOQGWKESDYAEILRQERDSTLTVASGPKADLRIRASGTIVEMUSRGQL  
KLKALRLAQGEYFRSQOQCLYLLDDFASLELTSRQLLAKLKSTQAQVFVSAI  
NPQITMDLGNKSMFRVENGKIEVQFQD"

3860. 6326

/gene="gyrB"

/locus\_tag="plu0004"

3860. 6274

/gene="gyrB"

/locus\_tag="plu0004"

/codon\_start=1

/transl\_table=11

/product="DNA gyrase subunit B"

/protein\_id="CAE12299.1"

/db\_xref="GI:36783459"

/translation="MSNTYDSSSIKVLGLDAVPRKPMYIGTDDTGTGHHMVFVV  
DNAIDALAGHCSEVITVTHADNSVSVQDDGRGPTGTHHEEGVSAEAVIMTVLHAGG  
KFDNSYKVSGLHGVGVVVALSEKELVIRRDGVKHEQTVHLGVQPSPFLVVGHT  
EQGTGRVAFWFSMDTFRDNTFEQHDILAKRLRELSFLNSGVSLRIDKRTNIEDHPHY  
EGGKAFVEFLNKMTPIHPNVFYFSTKDGIGVEISQWNGDFQENIYCFTNIPOR  
DGGTHLVGFRTAMRTILNSYMDKGVSKVSATGDDAREGLAVSVKVPKPS  
QTKDLVSEVKTAVETLMNEKLYELLENPDAKTVVSKI IDAARAREARKBMT  
RRKALDLAGLPGLADQERDPALSELVLEGGDSAGGSAKQGNENKQAILPLKSKI  
LNVEKARFDKMLSSQEVATILITAGCGIGEDVNPDKLRVHSIIMTDADVDSHIRT  
LLLPFYRQMPETIIEERGHVFIAPPLIKYKKGQEQYIKODEANDYFQUSIALDGASL  
VTNEHPALRGEPELEKLVTFENGQVKIIRKRLYPLSLNSLIYHPNLTETLSDKA  
KYENIEMLVTRLNDSQNSYSSYIHPENRERQMYEPVLCVTHGVDTYDYLDFDV  
HGSEYHRTLLDGLKRLNIEDAYIERGERQPVSPFEQALNLTKESSRGLSIORYK  
GLGEMNPQLWETTPETRRMRVTVKDAIATDQLFTTLMGDAVPRRAFTEENALK  
AANTDI"

terminator

6300. 6326

/gene="gyrB"

/locus\_tag="plu0004"

6415. 6926

/locus\_tag="plu0005"

6415. 6894

/locus\_tag="plu0005"

/function="Unknown"

/note="unnamed protein product; Similar to Hcp protein"

/codon\_start=1

/transl\_table=11

/protein\_id="CAE12300.1"

/db\_xref="GI:36783460"

/translation="MADMIYMTIKKKQGLISAGCSSVDSIGNKYQANHFDOILVYSL  
SHAITRTQNVHQPIIIQKPIDKSPLLGVAISENELSCSIDFRTNSNGAQERYN  
IKITGATITRISVLHPHSLNHNELPQESISLKYNITWNHLIAGTGGYSIWDRIY"

6908. 6926

/locus\_tag="plu0005"

/complement(8013..8778)

/locus\_tag="plu0006"

/complement(8013..8768)

/locus\_tag="plu0006"

/function="Unknown"

/note="unnamed protein product; Similar to unknown  
protein"

/codon\_start=1

/transl\_table=11

/protein\_id="CAE12301.1"

/db\_xref="GI:36783461"

/translation="MKQDNLDLKVRLELAKTLSFRVLTSHSKTRMDPNSKINLVE  
DILQVRIHCLSSNGAIEIIOPEIDHLKEQAFCLTANVKQYAIEREKESSTNLI  
LKOIFVGGGTQIILAGYTVCKASGLACASGAPLMAHGYNNVNGVYLLIYRENING  
GUREGVRYANKIGLSDDKADITYATVDLVLSGYGIFRKKVLKPREKSWLFRNINSDF  
IRGWRMGVSLTTEGIVDVSVLSIYQLEKEK"

complement(8773..8778)

/locus\_tag="plu0006"

/complement(8935..10056)

/gene="asd"

/locus\_tag="plu0007"

complement(8935..10044)

/gene="asd"

/locus\_tag="plu0007"

/transl\_table=11

/product="aspartate-semialdehyde dehydrogenase (ASA  
dehydrogenase) (ASADH)"

/protein\_id="CAE12302.1"

/db\_xref="GI:36783462"

/translation="MMKNVGVIGWRMGVSLMQRMIEQDFVIPCFFFTSQHGP  
APDFTGQGTQLQNAFDIIPALGADIIISCGGDYTNIEIYKPLKQWGYIDAAAL  
RWNDSIIILDPVNHTHIQGLNKGIKTFVGNCTVSLMLSLGGLFANDLVMSVA  
TYQASAGAGARHRELLVQMSLHTQVAKELQDPASAILDIERKVDTFRSGVMSTEK

```

FSVPLAGSLIPWIDKQNGSRBWKGAETNKILNTGNNTITVDGLCYRIGALRCH
SOAPTLKLDKDISIPEISOLLAHNDWVRVLPNDRESMRELTFAAVTGILDPVGL
RLNMGPEVLSAFTVGDQLWGAAEPLRMLRIIL"
complement(10051..10056)
/gene="asd"
/locus_tag="plu0007"
complement(10251..10472)
/gene="ogrK"
/locus_tag="plu0008"
complement(10251..10472)
/gene="ogrK"
/locus_tag="plu0008"
/function="Unknown, probable positive regulator of phage
P2 late gene transcription"
/codon_start=1
/transl_table=11
/product="Ogr protein"
/protein_id="CAE12303.1"
/db_xref="GI:36783463"
/translation="MIKPCGKAAHARSFHSCHTKERYNQCNINGCATFVSHT
FVTRISKPEVISVKPHKXTKTLQNV"
complement(10549..11634)
/locus_tag="plu0009"
complement(10549..11634)
/locus_tag="plu0009"
/function="Unknown, probable tail synthesis"
/note="unnamed protein product; Similar to protein D (gpd)
of Enterobacteria phage P2"
/codon_start=1
/transl_table=11
/protein_id="CAE12304.1"
/db_xref="GI:36783464"

Alignment Scores:
Pred. No.: 8.07e-72 Length: 349483
Score: 1004.00 Matches: 195
Percent Similarity: 84.08% Conservative: 48
Best Local Similarity: 67.47% Mismatches: 44
Query Match: 70.26% Indels: 2
DB: 1 Gaps: 1

US-09-545-199F-4 (1-289) x BX571859 (1-349483)

QY 1 MetAlaGlyAlaLysGluLeuThrLysIleAlaSerValLysSerThrGlnLysIle 20
Dd 37140 ATGCGCGCGCAAGAGATACGTACCAAGATCGCCAGCGTCAAAACACCAAAAATC 37081
QY 21 ThrLysAlaMetGluMetValAlaLysSerLysMetArgLysThrGlnGluArgMetSer 40
Dd 37080 ACTAAGCCGATGAGATGGTGTGCGTCCAAATGCGTAAACCGCAGATCGCATGCG 37021
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Dd 37020 GCCAGCGCTCTTATGCAGAAACCATACGACGCGATGTCACCTTGCCTTAGGPAAT 36961
QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Dd 36960 CTGGAAATAAACAATCCGTACCTTGAAGAGCGTCAAAACCAACGTCGCGTATCGTT 36901
QY 81 ValSerThrAspArgGlyLeuCysGlyGlyValLeuAsnValLeuPheLysThrValLeu 100
Dd 36900 GTTCTACGATCGTGGTGTGCGTGGTGTGATCTACTATCTGTTTCAAAAACATGCTT 36841
QY 101 ArgGluMetLysGluTyrPlyGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Dd 36840 TCAGAAATGAAGACTGTGTGATAAAGATCCAGTGTGAGTGGCGCTTATCGCTCT 36781
QY 121 LysSerIleAsnPheGlnSerLeuGlyLysIleLysIleLeuThrGlnAspSerGlyIle 140
Dd 36780 AAGCTACCTCTTCTTCTGTTCTGTTGGGGGATACGTTTCTCTCAGGTAACAGCATG 36721
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Dd 160

```

```

Db 36720 GGAGATAACCCCTTCACTGCTGTAATTGATCGGGCCAGTGAATATCATCTGCGGGGCATC 36661
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Dd 36660 GATGAGGACGTCGTGATAAACTGTATGTGTGATGATGATGATGATGATGATGATGATG 36601
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProLysLeuAspAsnAspGluLeuGly 200
Dd 36600 CAGGAACCGACTATTACTCAGTTATTACCTTCTGCTCCGAGATGATGAACACTG--- 36544
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Dd 36543 ---AAGAGAAATCTCGGATATTATACCAACCTGATCTTAAGCGTGTCTAGATATA 36487
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValValGluLeuLeuAlaSer 240
Dd 36486 CTGCTGCGTGTATGTAGATCGCAGTTATCAGGGCGTGTGATGATGATGATGATGATG 36427
QY 241 GluGlnAlaLysMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Dd 36426 GAACAGCCCGCAGAAATGTTGGCGATGAAAGCGCAACTGATATGTTGGTGGACCTGATC 36367
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Dd 36366 AAGAGTTGCGATGTTGTTTATACAAAGCTCGTCAGGCCAGCATAACTCAGGAGCTGACC 36307
QY 281 GluLeuValAlaGlyAlaAlaIle 289
Dd 36306 GAGATTGTCTCGGTGCTTCCGCGTT 36280
RESULT 14
AE004342/c
LOCUS
AE004342 12928 bp DNA linear BCT 10-OCT-2003
DEFINITION
Vibrio cholerae O1 biovar eltor str. N16961 chromosome I, section
250 of 251 of the complete chromosome.
ACCESSION
AE004342 AE003852
VERSION
AE004342.1 GI:9657358
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae O1 biovar eltor str. N16961
Vibrio cholerae O1 biovar eltor str. N16961
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
1 (bases 1 to 12928)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermlaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Uterback, T., Fleishmann, R.D.,
Nieman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL
Nature 406 (6795), 477-483 (2000)
MEDLINE
20406833
PUBMED
10952301
REFERENCE
2 (bases 1 to 12928)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermlaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Uterback, T., Fleishmann, R.D.,
Nieman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1. 12928
/organism="Vibrio cholerae O1 biovar eltor str. N16961"
/mol_type="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:243277"
/chromosome="I"
TITLE
JOURNAL
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source

```

```

/note="biotype: E1 Tor"
/136..2307
/gene="VC2758"
CDS
/136..2307
/gene="VC2758"
/note="similar to GB:M87049 SP:P21177 GB:M59368 GB:M74164
PID:145900; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="fatty oxidation complex, alpha subunit"
/protein_id="AAF95897.1"
/db_xref="GI:9657358"
/translation="MIYQKTIQVQLKQVLANGLAELSCAPASVKNLDELHLSLKDLD
ALAADSVKGLSSDEKAEIVGADITEFLGFAKEAELEDEWLFQANRINFKLELDL
FPTLSUKGTLGGCECVLATDFRIGDATTIGIPETKLMGPGFGVTLRLRIGA
DSAMEITQKACRAEALVGLLDAIVDSKLIDSATITLTOAIEEKLQKROOK
TSALTLSKLEAMNFTMAKGMVAQVAKHPAPMTSVVTIEAASLPDMDALDERKH
FKLAKSTEALVGLPLNDQYIKGLAKOSAKAASQDQTHRAVLGAGIMGGGIAYQSA
LKGVPVLMKDIAHSLLEGLWTEAAKLLNKQLERKIDGFKMAGILLASITPSLHVAGID
QADVIVBAVVENPKVAALVSEVEGLVDTEITLNTSTIPINLLAKSLKRPQFCGM
HFNPNHRELVLIIIGETSESDTINRVVAYAAKMGKSPFVYVNDPCGFFVNRVLPYF
AGSLMRDGANFTEIDKMERQFGFMGPAYLLDVGVIGIDTAHQAQVMAEGFFTRA
KSGREIDAILYEAKEFGQKNGSGFYQYTVDKKGPKAQFSDVLAILAPVCGAQPSD
POTLIERMIPMINEVVLGCEGIIASAQEDMALVYGLGPPRPGGVFRVLDITGIA
NYVMAEKYADLALYQVQLKQVLAQAQCTSFYSQAQYSAL"
/2333..3496
/gene="VC2759"
CDS
/2333..3496
/gene="VC2759"
/note="similar to GB:M87049 SP:P21151 GB:M59368 GB:M74164
PID:145901; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="fatty oxidation complex, beta subunit"
/protein_id="AAF95898.1"
/db_xref="GI:9657360"
/translation="MNTVVIVDCLTPMGRSKGAFRQRAEDLSAHLMKGILARNPQ
VNPKEIDYWCQVQTLQEQFNVARNAALAGLPIBIGAVTVNRLLGSSMWAALHDA
RAIMVGDABEICLVGGEVHGMVPMTHGVDFHPSKKNVAKAAGMGLTAMLSKLGHI
SRQODEPAARSHAAHATLEGPKNEILTEGHADGTLFTLDEIVIRPETHVAG
LAELRPDPANGTWTACTSALSADGASAMLVMSQKALGLTIRARIKMAVACD
PSIMYGVVPATHTKALQAGITQMDVVELNEFAAQSLPCAKDLGLEMMDDKYNL
NGGALAHGPGCSGSTRITSLINLMEAKDARYLATMCIQLGQGIATIFPR"
/complement(3563..4471)
/gene="VC2760"
CDS
/complement(3563..4471)
/gene="VC2760"
/note="similar to SP:P37598 GB:U00096 PID:1742732
PID:1742742 PID:1787949; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, LysR family"
/protein_id="AAF95899.1"
/db_xref="GI:9657361"
/translation="WMPKSKSLEMLDAVARLGSTFAAAVHLHKVPSAISYSVRQTESE
GLGVLPRLPKVELTIFAGELFMLEARQLLRQNEIRATQKAAHGWRTKLTVDLNV
VKLDKPMKMEAFQITDFALQINMEVNGSWEAIAQGRADIVIGATAVVPVGDFFE
VRDGLDWARFVMSPNPCVREQNLSGFTISQFLAICLDRTSSVLPKRHTHYPKORR
LLKPNWYSALSKSLGSLGVGMPRHMAQPLLAGSQVLEKVLDPDKLLSHCLLWREKDD
HKLIMWNVNGLGSDQLYSDWLQAK"
/4600..5829
/gene="VC2761"
CDS
/4600..5829
/gene="VC2761"
/note="similar to GB:U00096 PID:1742733 PID:1742743
PID:1787950; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="multidrug resistance protein"
/protein_id="AAF95900.1"
/db_xref="GI:9657362"
/translations="MNFPGHEKMKVSPFQLVYLSALSNLGFIAIDMYLPAFKQMETA
FATGPEQIALSLTVFLVGMALGQLLWGLSKDFEHRNTLIAGLIVFTLASAGLAPST
WQOLLRLRPTQAIQVCAPIVQAMVIRKPSVSCQLFATIMPLVALSPALPOLGV
LLADYFQWHSIETITLMGIVLVSTGUTNEIAEPTTS: AQDILKGLLKTLYLGNV
MWFACASAPFAITLTGMPETMAQGLGYQARDIGLSFIPOTIAPVGVGLKKAQVQYGD
ELVKQLIALFTLSAAMIFVASQWTLSSIPWILAPFCLIAVANGAAYPIVNRALSSA
KQSPATAAGLQNSLSQSSLSALVAAMASQAQVPTGIAVLCTGCTGLWGLVLSNRT
LREQFTTPNSRVSDSE"
/complement(5891..7252)
/gene="VC2762"
CDS
/complement(5891..7252)
/gene="VC2762"
/note="similar to SP:P17114 PID:290578 PID:43267
PID:290579 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine pyrophosphorylase"
/protein_id="AAF95901.1"
/db_xref="GI:9657363"
/translation="MKFSVTILAAKQTRMHSNMPKVLHTLAGKPMVKHVHIDTCNNLG
ADNHLVTHGGDQMQALANENVNVLQQLGTGHAVDAQSPHFODDEKILVLXGD
VELISEDITELLEAQPTDGIALLTVLEPTGYGRIVRKGPPVAIVEOKDASEROK
LIKEYNTGLVATGRLKRWLAGLNNNAQGEYVLTQVIAAHDEGSAVZAPHSBSI
EYGVNDRIQLARERAFQARQAKLLEQGMLEDPARFLRGLTQCGSDVEIDVNVV
IEGNVIGNNVVIGAGSILKDCEDDNTVIRPSIEGATVGENCTVPTFLRPGAE
LRDDAHVGNFVEMKNALGEGSKANHLTYLGDAIGKGVNVGAGVITCNPDGANKKT
VTGDDVFVSGDCQLVAPVTIGNGATIGAGTTLTKNVAEGLVITRAPERKIAGWQRP
AAA"
/complement(7499..7921)
/gene="VC2763"
CDS
/complement(7499..7921)
/gene="VC2763"
/note="similar to SP:P2988 PID:48340; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, epsilon subunit"
/protein_id="AAF95902.1"
/db_xref="GI:9657364"
/translations="MAAIFHLVDVSAEKKLSGLVETPQVTGSEGEGLIPHGHTPLL
TAKPGMVRIVKLHGHEEIVYSGVIEVQPGTATVLADTAIRGEDLDAKAEAKRR
ABERILNQHDMQFAQAASELAKAIAQLRVIELTKRR"
/complement(7940..9343)
/gene="VC2764"
CDS
/complement(7940..9343)
/gene="VC2764"
/note="similar to GB:X16050 SP:P12986 PID:48339;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, beta subunit"
/protein_id="AAF95903.1"
/db_xref="GI:9657365"
/translation="NATGKIVQIICAVVDVBPQSEVSVVDALNVDSKERLVLEVO
QQLGGVIRAIWVGSSDGLRGMVYQNTGAPISVPGTKTLGRIMNVLGDAIDERGDI
GAEEVYIHRPAPSYEESSATLELTGVKIDILCPFAKGGKILGFGGAGVKTVMV
MELINNALVHSGLSVFAVGERTREGNDHFHEMGAGVNVNVEQPSALLGKMSAVGQV
EPFNKRALVALTMAEKFEDEGRDVLFFDNIYRYTLAGTEVSALLGKMSAVGQV
PILAEEMGLQERTSTKKGITSVQAVVPADLTDPSPTTFAHLDATVVLNRTA
AMGLYPAIDPLDST"
/complement(9378..10244)
/gene="VC2765"
CDS
/complement(9378..10244)
/gene="VC2765"
/note="similar to SP:P12990 PID:48338; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, gamma subunit"
/protein_id="AAF95904.1"
/db_xref="GI:9657366"

```

```
/translation="MAGAKBIRTKIGSVKSTOKITKAMENVAASVRRSQDAMESRRP
YAOTIRKIVGHVANASLEVPYLERBBAKRVYIIISTDRGLCGGLNINLFKAITD
MQTKWEKGAQIELAIIIGSKATAFNNAGAAQVSGDLGSPSLDLTGSVGMILKXY
DGLGDLRLYLTFNGFVNTWOKPKIDOLLPLPKSDSEDMQRDHMDYIYEPEPKLLD
ALLRFLFESQVYQGVENACEQAARVAMKAATDNASNLIDDLQLVYNKARQAATQ
ELSEIVGGAAGAAV
complement(10289..11860)
/gene="VC2766"
CDS
complement(10289..11860)
/gene="VC2766"
/notes="similar to SP:PI2985 PFD:48337; identified by
sequence similarity; putative"
/codon_start=1
```

## Alignment Scores:

Pred. No.:	3.94e-73	Length:	13928
Score:	998.50	Matches:	188
Percent Similarity:	83.74%	Conservative:	54
Best Local Similarity:	65.05%	Mismatches:	46
Query Match:	69.87%	Indels:	1
DB:	1	Gaps:	1

US-09-545-199F-4 (1-289) x AE004342 (1-12928)

QY	1	MetAlaGlyAlaLysGluLeuArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
Db	10244	ATGCCCGCGCAAAAGAGATACGTACTAAGTCGGAGGTGTAAGAGCACTCAGAAGATT 10185
QY	21	ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
Db	10184	ACGAAAGCGATGGAATGTGGTCTGTTCAAAATATGCGTCGTTCACAGAGCGTATGGAG 10125
QY	41	SerSerArgProTyrSerGluThrIleArgAenValIleSerHisValSerLysAlaThr 60
Db	10124	TCATCAGCGCCCTTACCAAAACAATACGTAAAGTATGATCGGTGTCGCCAACGCAAGT 10065
QY	61	IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Db	10064	CTAGATATCGTATCCGTTATCTAGAGAGCGGTGAAGCGGAGCGAGTGGTTATCATATT 10005
QY	81	ValSerThrAspArgGlyLeuGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db	10004	ATTCTACAGACCGCGGTTGTGTGGTGTGTGAACATTATCTGTTCAAAAAGCCATC 9945
QY	101	AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db	9944	ACGGATATGCCAACTTGGAAAGAGAAAGTCGCGCAGATTGAGCTGGCGGATATCGGCTCG 9885
QY	121	LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db	9884	AAAGCGACCGCTTTTTCACAAACAGCGCGCGAAAGTAGCGGCTCAGGTTTCTGTTTG 9825
QY	141	GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db	9824	GCGCACGCCCAAGCTTAGAAGATCTGATCGGTTGCGTTCGTCATCGTGAAGAATAC 9765
QY	161	LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db	9764	GATAAAGGTGATGATAGACCGCTCTATCTGTGTTCACACAGTTTGAACACCATGGTG 9705
QY	181	GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAsnAspGluLeuGly 200
Db	9704	CAAAACCAAGATCGATCAATCTACCTTTGCAAAATCGGACGCGAAGATATG--- 9648
QY	201	GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db	9647	CAGCGTGATCATGTGGGACTACATCTACGAGCGTCAACCAAGGCTCTATTAGATGCG 9588
QY	221	LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240
Db	9587	CTTTTATCGGCTTCATGTAGTGCAGGTTTATCAAGGTTTATCAAAAACCTAGCGTGT 9528
QY	241	GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeulle 260

Db	9527	GAGCAACGACCCGAATGGTTGGATGAAGCTGTACCGATACCGGAGCAACCTGATT 9468
QY	261	AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db	9467	GATGATTTCAACTTGTGTACAAACAAAGCCGTCAGGCGCGGATTCACACAAATTTGTCG 9408
QY	281	GluIleValAlaGlyAlaAlaIle 289
Db	9407	GAAATCGTAGGCGCGCGCTGCGGTT 9381

## RESULT 15

BX950851.50/c

## WPCOMMENT

Sequence split into 51 fragments LOCUS BX950851 Accession BX950851

Fragment Name	Begin	End
BX950851_00	1	110000
BX950851_01	100001	210000
BX950851_02	200001	310000
BX950851_03	300001	410000
BX950851_04	400001	510000
BX950851_05	500001	610000
BX950851_06	600001	710000
BX950851_07	700001	810000
BX950851_08	800001	910000
BX950851_09	900001	1010000
BX950851_10	1000001	1110000
BX950851_11	1100001	1210000
BX950851_12	1200001	1310000
BX950851_13	1300001	1410000
BX950851_14	1400001	1510000
BX950851_15	1500001	1610000
BX950851_16	1600001	1710000
BX950851_17	1700001	1810000
BX950851_18	1800001	1910000
BX950851_19	1900001	2010000
BX950851_20	2000001	2110000
BX950851_21	2100001	2210000
BX950851_22	2200001	2310000
BX950851_23	2300001	2410000
BX950851_24	2400001	2510000
BX950851_25	2500001	2610000
BX950851_26	2600001	2710000
BX950851_27	2700001	2800001
BX950851_28	2800001	2910000
BX950851_29	2900001	3010000
BX950851_30	3000001	3110000
BX950851_31	3100001	3210000
BX950851_32	3200001	3310000
BX950851_33	3300001	3410000
BX950851_34	3400001	3510000
BX950851_35	3500001	3610000
BX950851_36	3600001	3710000
BX950851_37	3700001	3810000
BX950851_38	3800001	3910000
BX950851_39	3900001	4010000
BX950851_40	4000001	4110000
BX950851_41	4100001	4210000
BX950851_42	4200001	4310000
BX950851_43	4300001	4410000
BX950851_44	4400001	4510000
BX950851_45	4500001	4610000
BX950851_46	4600001	4710000
BX950851_47	4700001	4810000
BX950851_48	4800001	4910000
BX950851_49	4900001	5010000
BX950851_50	5000001	5064019

Continuation (51 of 51) of BX950851 from base 5000001 (BX950851 Erwinia carotovora subsp.

## Alignment Scores:

Pred. No.:	8.22e-72	Length:	64019
Score:	993.00	Matches:	194
Percent Similarity:	82.70%	Conservative:	45

Best Local Similarity:	67.13%	Mismatches:	48
Query Match:	69.49%	Indels:	2
DB:	1	Gaps:	1
US-09-545-199F-4 (1-289) x BX950851_50 (1-64019)			
QY	1	MetAlaGlyAlaLysGluLilAArgThriLysIleAlaSerValLysSerThrGlnLysIle	20
DB	56259	ATGGCGCGCGCAAAAGATACGTAGTAGATCCAGCGTCCAAAATACGCAGAAGATC	56200
QY	21	ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer	40
DB	56199	ACCAAAGCAATGGAAATGGTCGCCGCTCTCAAAATGCGTAATACCGAGGAGGCTATGGCG	56140
QY	41	SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr	60
DB	56139	GCACCGCTCTTATGGCGAAACATACGCAATGTGATTGGTCACCTGCGTTAGGGAAT	56080
QY	61	IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal	80
DB	56079	CTGGAATATAAACACCCGTACCTGGAAGACGTGGCGTTAAGCGGCTTGGGTATTGGTG	56020
QY	81	ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLysPheLysThrValLeu	100
DB	56019	GTGTCTACTAGCGTGGCTGTGTGTGTGTGTGAACATTAACTCTTCAAAAACACTGCTG	55960
QY	101	AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer	120
DB	55959	GCTGATATGAATCCTTGGACGCAAGCGGTGAACTGATTAGCCCTGATTGGTTCC	55900
QY	121	LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle	140
DB	55999	AAAGCGCTTCTTCTTCGTGTCGTGAGCGGAAACATTGTTGCTCAGGTTACCGGTATG	55840
QY	141	GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr	160
DB	55839	GGGGATAACCTTCCGTATCAGAAATGATCGGCCCGGTTAAAGTTATGCTGCAAGCCAT	55780
QY	161	LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer	180
DB	55779	GACGAAGGTGCTGTGACAAAGCTGATATCGTAAGCAACAAGTTTATCAATACCATGCT	55720
QY	181	GlnLysProValLeuGlnLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly	200
DB	55719	CAGGAACCGCTTATTCTTCAAGTGTACCGTTACGCCCTTCAGATACAGTAGGTG---	55663
QY	201	GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn	220
DB	55662	---AAGAAAGAACCTGGGAATACCTGTATGAACCCGACTCTAAGTCGCTGCTGGATACC	55606
QY	221	LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer	240
DB	55605	CTGCTCGCGGTATGTAGATCTCAGGTTTATCAGGGGCTGTAGAAAATCTGGCTAGT	55546
QY	241	GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle	260
DB	55545	GAGCAGCGCGCGAATGGTTGCGATGAAGCCCGACTCATACCGCGTAGCTGATC	55486
QY	261	AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn	280
DB	55485	AAAGAGCTTCAGTTGGTATACACNAGCGCGCTCAGCCAGCAATTACTCAGAACTACC	55426
QY	281	GluIleValAlaGlyAlaAlaAlaIle	289
DB	55425	GAATTCGTGGGGAGCGCTCGCGGCT	55399

Search completed: November 9, 2004, 14:48:48  
Job time : 24192 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 9, 2004, 02:54:00 ; Search time 629 Seconds  
(without alignments)  
2411.896 Million cell updates/sec

Title: US-09-545-199F-4

Perfect score: 1429

Sequence: 1 MAGAKEIRTKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapext 0.5			
Delop 6.0			7.0
Delext 7.0			

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2/1/US070.spool\_p/US09545199/runat\_08112004\_112526\_20941/app\_query.fasta\_1.455  
-DB=N\_Geneseq\_238sep04 -Qfmt=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOPCPL=0  
-LOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=pt0 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09545199 @CGN 1.1.885 @runat\_08112004\_112526\_20941 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

	N_Geneseq_238sep04.*
1:	Geneseqn1980s.*
2:	Geneseqn1990s.*
3:	Geneseqn2000s.*
4:	Geneseqn2001as.*
5:	Geneseqn2001bs.*
6:	Geneseqn2002as.*
7:	Geneseqn2002bs.*
8:	Geneseqn2003as.*
9:	Geneseqn2003bs.*
10:	Geneseqn2003cs.*
11:	Geneseqn2003ds.*
12:	Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1429	100.0	1972	3 AAC79582	Aac79582 Virulence
2	1429	100.0	1972	6 ABQ83459	Abq83459 Pasteurel
C 3	1173	82.1	110000	2 AAT42063_04	Continuation (5 of
C 4	1173	82.1	110000	2 AAT42063_05	Continuation (6 of
5	1132.5	79.3	867	3 AAC79654	Aac79654 Virulence
6	1132.5	79.3	867	6 ABQ83531	Abq83531 Actinobac

7	1077.5	75.4	866	6 ABQ83549	Abq83549 Pasteurel
8	1004	70.3	864	10 ACP70377	Adf70377 Photorhab
9	1004	70.3	876	10 ADF02992	Adf02992 Bacterial
10	1004	70.3	110000	10 ACF65385_5	Continuation (6 of
11	1004	70.3	110000	10 ACF67367_34	Continuation (35 o
12	991	69.3	864	5 AAH81345	Aah81345 Escherich
13	978	68.4	894	11 ACH97650	Ach97650 Klebsiell
14	926.5	64.8	876	11 ABD03905	Abd03905 Pseudomon
C 15	926.5	64.8	1050	11 ABD03901	Abd03901 Pseudomon
16	894	62.6	879	9 ADA29481	Ada29481 DNA encod
17	885	61.9	930	12 ADL03242	Adl03242 DNA encod
18	885	61.9	94750	4 AAF28551	Aaf28551 Genomic f
19	804	56.3	110000	6 ABA32787_0	Abas32787 Buchnera
20	760	53.2	873	10 ABZ38350	Abz38350 N. gonorr
C 21	753	52.7	46593	3 AAA81456	Aaa81456 N. mening
C 22	753	52.7	110000	3 AAA81489	Continuation (6 of
C 23	753	52.7	349980	3 AAF21612_5	Aaf21612 Neisseria
C 24	750	52.5	3125	5 AAS78013	Aas78013 DNA encod
25	750	52.5	3125	5 AAS88668	Aas88668 DNA encod
26	739	51.7	3936	5 AAS82564	Aas82564 DNA encod
27	702	49.1	2786	5 AAS93030	Aas93030 DNA encod
28	673	47.1	3091	5 AAS82566	Aas82566 DNA encod
29	650.5	45.5	531	3 AAC79635	Aac79635 Virulence
30	650.5	45.5	531	6 ABO83512	Abg83512 Actinobac
31	616	43.1	2583	5 AAS89854	Aas89854 DNA encod
32	530.5	37.1	915	10 ADH84612	Adh84612 Enterococ
33	530.5	37.1	7750	2 AAX13153	Aax13153 Enterococ
34	530.5	37.1	7750	6 ABS98948	Abs98948 Enterococ
35	524.5	36.7	4815	2 AAV29571	Aav29571 L. lactis
36	523.5	36.6	110000	6 ABA90521_18	Continuation (19 o
C 37	518.5	36.3	879	6 ABN68209_08	Abn68209 Streptoco
38	518.5	36.3	110000	6 ABN71527_08	Continuation (9 of
39	518	36.2	851	6 ABK72809	Abk72809 Bacillus
40	510.5	35.7	903	10 ADC94003	Adc94003 E. faeciu
41	508.5	35.6	873	6 ABN68210	Abn68210 Streptoco
42	504.5	35.3	2207	2 AAV29572	Aav29572 L. lactis
43	501.5	35.1	876	10 ABX07252	Abx07252 S. pneumo
44	501.5	35.1	888	8 ABZ42460	Abz42460 Streptoco
45	501.5	35.1	10399	2 AAV52293	Aav52293 Streptoco

## ALIGNMENTS

RESULT 1

AAC79582  
ID AAC79582 standard; DNA; 1972 BP.

XX  
AAC79582;

XX  
08-FEB-2001 (first entry)

XX  
Virulence gene #2.

XX  
Virulence gene; antibacterial; vaccine; bacterial infection; septicemia;

XX  
bronchopneumonia; rhinitis; wound infection; ss.

XX  
Pasteurella multocida.

XX  
WO2000061724-A2.

XX  
19-OCT-2000.

XX  
06-APR-2000; 2000WO-US009218.

XX  
09-APR-1999; 99US-0128689P.

XX  
10-SEP-1999; 99US-0153453P.

XX  
(PHAA ) PHARMACIA & UPJOHN INC.

XX  
Lowery DE, Fuller TE, Kennedy MJ;

XX  
WPI; 2000-647422/62.

XX  
P-FSDB; AAB44522.

XX Attenuated Pasteurellaceae bacteria comprising mutations in virulence  
PT genes, useful as a live attenuated vaccine against bacterial infections.  
XX  
PS Claim 1; Page 67-68; 322pp; English.  
XX  
CC The family Pasteurellaceae encompasses several pathogens that infect a  
CC wide variety of animals. The present invention relates to virulence genes  
CC from Pasteurellaceae. The present sequence is one such virulence gene.  
CC The present sequence may be mutated in order to produce an inactivated gene.  
CC The inactivated virulence gene may in turn be used to produce a vaccine,  
CC which is useful for treating bacterial infections such as septicemias,  
CC bronchopneumonias, rhinitis and wound infections  
XX  
SQ Sequence 1972 BP; 652 A; 312 C; 438 G; 570 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1-88e-144 Length: 1972  
Score: 1429.00 Matches: 289  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-545-199F-4 (1-289) x AAC79582 (1-1972)

QY 1 MetAlaGlyValLysGluLeuThrLysLeuAlaSerValLysSerThrGlnLysLeu 20  
Db 364 ATGGCAGGTCCTAAAGAGATAGAACCAAAATCGCGAGTGTAAGAAAGTACACAAAAATT 423  
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysLysMetArgLysThrGlnLysArgMetSer 40  
Db 424 ACTAAGCGATGGAATGGTTCCTCGAATGCGTAAGCAAGCAAGCATGTCT 483  
QY 41 SerSerArgProTyrSerGluThrLysLeuArgSerHisValSerLysAlaThr 60  
Db 484 TCTTCACGCCCTTATTCAGAACCAATACGTAACTGATTAGCCAGCTTCCCAAGCAAG 543  
QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetLeuVal 80  
Db 544 ATGGTTACAGCATCCATTTTGTAGTGATCGCAAGTAAAGAAAGTGGCGATGTT 603  
QY 81 ValSerThrAspArgGlyLeuCysGlyLeuLysValAsnValAsnLeuPheLysThrValLeu 100  
Db 604 GTGTCCACAGATCGTGTCTTTGTGGTCTTAAACGTGAATTTTAACTGTATTA 663  
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLysSerLeuLeuGlySer 120  
Db 664 AATGAATGAAGATGGAAGAAAGAAAGATTTTCGTTCAATTTGATTTAATCGTTCT 723  
QY 121 LysSerLeuAsnPheGlnSerLeuGlyLysLeuLysLeuThrGlnAspSerGlyLeu 140  
Db 724 AAATCTATCAACTTTTCCCAATCTTTGGGAATTTAAATTTTAAACCAAGATTCAGTATT 783  
QY 141 GlyAspThrProSerValGluGlnLeuLysGlySerValAsnSerMetIleAspAlaTyr 160  
Db 784 GGTAATCTCCCTCTGTTGACGAGTAAATGGTTCAGTCAATTTCTATGATTCAT 843  
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
Db 844 AAAAAAGGGAAGTAGATGTTGTATTAGTTTATACAAATTTATTAACCAAGATTCG 903  
QY 181 GlnLysProValLeuGluLysLeuLeuProLeuProGluLeuAspAsnAspGluLeuGly 200  
Db 904 CAAGACCGATTTGGAAAAATTAATTCATPACCAGAAATTAGATAATGATGATTAGGC 963  
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
Db 964 GAAAGAAAAACAAGTTGGGATTTATTTACGACCTGATGCGAAGATTTATTAGATAAT 1023  
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuLeuAsn 240  
Db 1024 TTATTGGTTTCGTTATTAGATCTCAGGTTTATCAAGCAGCAGTTGAAACCTTGCTTCT 1083

QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLeu 260  
Db 1084 GACCAAGCCGCTCGAATGCTGCCATGAAGCAGCAACAGATAAGCCAGGTAACTTAATT 1143  
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
Db 1144 AATGAGTTACAGTTAGTCTATTAACCAAGCTGCTAAGCAAGTATTACCAATGAATTAAT 1203  
QY 281 GluLeuValAlaGlyAlaAlaAlaLeu 289  
Db 1204 GAATTTGTCGCGTGCACGACCAATT 1230

RESULT 2

ABQ83459  
ID ABQ83459 standard; DNA; 1972 BP.  
XX  
AC ABQ83459;  
XX  
DT 07-AUG-2003 (revised)  
DT 24-JAN-2003 (first entry)  
XX  
DE Pasteurella multocida atpG gene SEQ ID NO:3.  
XX  
KW Antibacterial; vaccine; gram negative bacterial virulence gene;  
KW identification; virulence; Pasteurellaceae; gene; ds.  
XX  
OS Pasteurella multocida.  
XX  
FN WO200275507-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 17-JAN-2002; 2002WO-US0001971.  
XX  
PR 15-MAR-2001; 2001US-00809665.  
XX  
PA (PHEA) PHARMACIA & UPJOHN CO.  
XX  
PI Lowery DE, Fuller TE, Kennedy MJ;  
XX  
DR WPI; 2002-740866/80.  
DR P-PSDB; ABP54474.  
XX  
PT New mutant gram-negative bacteria, useful as vaccines and for identifying  
PT new anti-bacterial agents that target virulence genes and their products.  
XX  
PS Claim 1; Page 72-73; 350pp; English.  
XX

CC The present invention describes a gram-negative bacteria comprising a  
CC mutation in a gene, where the mutation results in decreased activity of a  
CC gene product encoded by the mutated gene. Also described is a method for  
CC producing a gram-negative bacteria mutant or an attenuated  
CC Pasteurellaceae bacteria. The mutated genes have antibacterial activity  
CC and can be used in vaccines. The gram-negative bacteria or the attenuated  
CC Pasteurellaceae bacteria can be used as vaccines in the fields of human  
CC medicine or veterinary medicine, and for identifying new antibacterial  
CC agents that target the virulence genes and their products. ABQ83458 to  
CC ABQ83578 and ABP54473 to ABP54551 represents sequences used in the  
CC exemplification of the present invention. (Updated on 07-AUG-2003 to  
CC correct OS field.)  
XX

SQ Sequence 1972 BP; 652 A; 312 C; 438 G; 570 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1-88e-144 Length: 1972  
Score: 1429.00 Matches: 289  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-545-199F-4 (1-289) x ABQ83459 (1-1972)

QY 1 MetAlaGlyAlaLysGluLeuArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
Db ATGGCAGGTGCTAAAGAGATAGAACCAAAATCGAGTGTAAAAAGTACACAAAAAT 423  
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
Db ACTAAGCGATGGAANTGGTCTGCTCCGAAATGCGTAAAAACGCAAGACGATGTCT 483  
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
Db TCCTTACGCCCTTATTAGAAACAAATACGTAACGTGATTAGCCACGTTTCCAAAGCAACG 543  
QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
Db ATGGTTTACAAGCATCCATTTTAGTGCGATCGGAAGTAAAAAAGTGGCGATGATTGTT 603  
QY 81 ValSerThrAspArgGlyLeuCysGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
Db GTGTCCACAGATCGTGTCTTTGTGGTGGCTTAAACGTGAACCTGTTTAAAACTGTATTA 663  
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120  
Db AATGAATGAAGAATGGAAGAAAGAGATGTTTCGTTCAATTGAGTTTAAATCGTTCT 723  
QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140  
Db AATCTATCAACTTTTCCAACTTTTGGCAATTAAATTTTAAACCAAGATTCAGGTATT 783  
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaThr 160  
Db GGTGATACTCCCTCTGTTGAGCAGTTTAAATGTTTATTAACAAATTTATTAACAGATGTCG 843  
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
Db AAAAAAGGGAAGTAGATGTTGTTATTAGTTTATTAACAAATTTATTAACAGATGTCG 903  
QY 181 GlnLysProValLeuGlnLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
Db CAAAGCCAGTATTGGAAAAATTAATTCATTAACAGAAATAGATAATGATTAAT 963  
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAsn 220  
Db GAAAGAAACAAAGTTGGGATTATTTACCACTGATCGGAAGTATTATTAGATAAT 1023  
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGlnLeuLeuAlaSer 240  
Db TTTTGGTTCGTTATTAGAAATCTCAGTTTATCAAGCAGCTTGAAACCTTGCTTCT 1083  
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260  
Db GAGCAAGCCGCTCGAATGGTCCCATGAAAGCAGCAACAGATAACGAGTAACTTAAT 1143  
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
Db AATGAGTTACAGTTAGTCTATAACAAAGCTCGTCAAGCAAGTATTACAAATGAATTAAT 1203  
QY 281 GluIleValAlaGlyAlaAlaIle 289  
Db GAAATTCGTGCGGTGCGAGCGCAAT 1230  
RESULT 3  
AAT42063\_04/c  
Continuation (5 of 19) of AAT42063 from base 400001 (Haemophilus influenzae complete gen  
p Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063  
p Fragment Name Begin End  
p AAT42063\_01 1 110000  
p AAT42063\_02 100001 210000  
p AAT42063\_03 200001 310000  
p AAT42063\_04 300001 410000  
p AAT42063\_05 400001 510000  
p AAT42063\_06 500001 610000  
p AAT42063\_07 600001 710000

WP AAT42063\_07 700001 810000  
WP AAT42063\_08 800001 910000  
WP AAT42063\_09 900001 1010000  
WP AAT42063\_10 1000001 1110000  
WP AAT42063\_11 1100001 1210000  
WP AAT42063\_12 1200001 1310000  
WP AAT42063\_13 1300001 1410000  
WP AAT42063\_14 1400001 1510000  
WP AAT42063\_15 1500001 1610000  
WP AAT42063\_16 1600001 1710000  
WP AAT42063\_17 1700001 1810000  
WP AAT42063\_18 1800001 1930121  
Alignment Scores:  
Pred. No.: 2,31e-114 Length: 110000  
Score: 1173.00 Matches: 221  
Percent Similarity: 91.00% Conservative: 42  
Best Local Similarity: 76.47% Mismatches: 26  
Query Match: 82.09% Indels: 0  
DB: 2 Gaps: 0  
US-09-545-199f-4 (1-289) x AAT42063\_04 (1-110000)  
QY 1 MetAlaGlyAlaLysGluLeuArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
Db 101934 ATGGCAGGTGCAAAAGAGATAAAAACCAAAATTCGAGTGTACAAAGTACACAAAAATC 101875  
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
Db 101874 ACTAAGCGAATGGAATGTTGGCAACCTCGAAATTCGTAACCGCAGGATCGTATCGCT 101815  
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
Db 101814 GCATCTCGTGTATTCGAACTATCCGTAACTATTAGTTCATGTGTAAAGGCAAGT 101755  
QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
Db 101754 ATCGTTTATAAACATCCGTTCTTAGTTAGCGGCAAGTGAAGAAATCGTATCTTTGGTT 101695  
QY 81 ValSerThrAspArgGlyLeuCysGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
Db 101694 ATTCAACAGATCGTGGATGTTGGTGGTTAAATGTTAAATTTATTAATCAAAACCACTT 101635  
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120  
Db 101634 ACCAAATAAAAATTCGAAAGAACAAATATTCTACAGATTGGGCTTAATAGTTCA 101575  
QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140  
Db 101574 AAAGGATTAGTTTTTTCGTTCTTGGATTAAATATCAAAAGTCAGCTTTCTCGTTTA 101515  
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaThr 160  
Db 101514 GCGATACGCCGCTCTAGAAATTAATTTGGTGGCAATACAAATGTTGATGCTTAT 101455  
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
Db 101454 CGTAATGTTGAATGTGTCAGTTTATTTGATGATCAATAAATTTGTTAAATACGATGTCG 101395  
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
Db 101394 CAAAAGCTGTTGTACAAACATTAAGTCTTACCAAGTAACTAAAGCAGATCAATTTAAAT 101335  
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
Db 101334 GAAAGCAACAGACTTGGGATTAATTTATGAGCGAGAACCAAAAGTACTATTAGATAGC 101275  
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGlnLeuLeuAlaSer 240  
Db 101274 CTTTATGTTCTGTTATTAGAGTCCCAAAATTTATCAAGCGTTGTAGATGATGATGTTCA 101215  
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260



Db 101214 GAACACAGCGGCTCGAATGTTAGCAATGAACAGCAGCAACGGATAATCGAGTTAATTAAATT 101155

Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280

Db 101154 AATGATCTGGGTGGTGTACACAAAGCTCGTCAAGCAAGTATCACAAATGAATTGAAT 101095

Qy 281 GluIleValAlaGlyAlaAlaIle 289

Db 101094 GAAATCGTAGCGGTGGGGCGGATT 101068

RESULT 4

AAT42063\_05/c

Conjunction (6 of 19) of AAT42063 from base 500001 (Haemophilus influenzae complete genome sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063

WP	Fragment Name	Begin	End
WP	AAT42063_00	1	110000
WP	AAT42063_01	100001	210000
WP	AAT42063_02	200001	310000
WP	AAT42063_03	300001	410000
WP	AAT42063_04	400001	510000
WP	AAT42063_05	500001	610000
WP	AAT42063_06	600001	710000
WP	AAT42063_07	700001	810000
WP	AAT42063_08	800001	910000
WP	AAT42063_09	900001	1010000
WP	AAT42063_10	1000001	1100000
WP	AAT42063_11	1100001	1210000
WP	AAT42063_12	1200001	1310000
WP	AAT42063_13	1300001	1410000
WP	AAT42063_14	1400001	1510000
WP	AAT42063_15	1500001	1610000
WP	AAT42063_16	1600001	1710000
WP	AAT42063_17	1700001	1810000
WP	AAT42063_18	1800001	1830121

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
2.31e-114	1173.00	110000	221
Percent Similarity:	91.00%	Conservative:	42
Best Local Similarity:	76.47%	Mismatches:	26
Query Match:	82.09%	Indels:	0
DB:	2	Gaps:	0

US-09-545-199F-4 (1-289) x AAT42063\_05 (1-110000)

Qy 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20

Db 1934 ATGCGAGTGCACAAAGAGATTAACAAACAAATTCGCCAGTGTACAAAGTACACAAATATC 1875

Qy 21 ThrLysAlaMetGluMetValAlaIleAlaSerLysMetArgLysThrGlnLysArgMetSer 40

Db 1874 ACTAAGGCAATGGAATGTTGGTGGCACTCGAATAATGCGTAAACCCAGATCGTATGCT 1815

Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60

Db 1814 GCATCTGTCGCGTATCTGAAACTATCCGTAACGTTATAGTATAGTATAGTATAGTATAGT 1755

Qy 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal 80

Db 1754 ATCGGTATATAACATCCGTTCTTATTGTCGCGGAGTGAAGAAATTCGGTATCTCGTT 1695

Qy 81 ValSerThrAspArgLysLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100

Db 1694 ATTTCAACAGATCGGGGATGTTGGTGGTAAATGTTAATTTATTTCAAAACCACTT 1635

Qy 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120

Db 1634 AACCAAAATATAAAATTTGGAAGACAAATATTTCTACAGATTTGGGCTTAATAGGTTC 1575

Qy 121 LysSerIleAsnPhaPheGlnSerLeuGlyLysIleLeuThrGlnAspSerGlyIle 140

Db 1574 AAAGGATAGTTTTCCTGCTTCTTGGATTAATCAAGGTCAGCTTCTGTTTGA 1515

Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160

Db 1514 GGGATACCCCGCTCTAGAGAAATTAATGTTGGCAATACAAATCTTTGATGCTTAT 1455

Qy 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180

Db 1454 CGTAATGTTGAATGATGCTGATTTATTTATTCACAAATAAATTTGTTAATACGATGTC 1395

Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnGluLeuGly 200

Db 1394 CAAAAGCCTGTTGTACAAATTAATTCCTTTACCAGAATCTAAAGACGATCATTTAAAT 1335

Qy 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220

Db 1334 GAAAGACACAGACTTGGGATTTCTTTATGAGCAGACAAAGTACTATTAGATAGC 1275

Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240

Db 1274 CTTTATGTTCTGTTATTTAGAGTCCCAATTTATCAAGCGTTGTAGATATGTAGTTCA 1215

Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260

Db 1214 GAACAGCGCTCGAATGTTAGCAATGAAAGCAGCAACGATAATCGAGTAATTTAAT 1155

Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280

Db 1154 AATGATCTGGGTTGGTGTACAAACAAAGCTCGTCAAGCAAGTATCAAAATGAATTGAAT 1095

Qy 281 GluIleValAlaGlyAlaAlaIle 289

Db 1094 GAAATCGTAGCGGTGGGGCGGATT 1068

RESULT 5

AAC79654

ID AAC79654 standard; DNA; 867 BP.

XX AAC79654;

AC AAC79654;

DT 08-FEB-2001 (first entry)

XX Virulence gene #61.

XX Virulence gene; antibacterial; vaccine; bacterial infection; septicemia;

XX Virulence gene; antibacterial; vaccine; bacterial infection; septicemia; bronchopneumonia; rhinitis; wound infection; ss.

XX Actinobacillus pleuropneumoniae.

XX WO200061724-A2.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US009218.

XX 09-APR-1999; 99US-0128689P.

XX 10-SEP-1999; 99US-0153453P.

XX (PHAA) PHARMACIA &amp; UPJOHN INC.

XX Lowery DE, Fuller TE, Kennedy MJ;

XX WPI: 2000-647422/62.

XX P-PSDB; AAB44579.

XX Attenuated Pasteurellaceae bacteria comprising mutations in virulence genes, useful as a live attenuated vaccine against bacterial infections.

XX Claim 7; Page 290-291; 322pp; English.

XX The family Pasteurellaceae encompasses several pathogens that infect a wide variety of animals. The present invention relates to virulence genes from Pasteurellaceae. The present sequence is one such virulence gene. The present sequence may be mutated in order to produce an inactive gene. The inactive virulence gene may in turn be used to produce a vaccine,

CC which is useful for treating bacterial infections such as septicemias,  
 CC bronchopneumonias, rhinitis and wound infections  
 XX  
 SQ Sequence 867 BP; 292 A; 147 C; 197 G; 231 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 7,71e-113 Length: 867  
 Score: 1132.50 Matches: 215  
 Percent Similarity: 89.97% Conservativity: 45  
 Best Local Similarity: 74.39% Mismatches: 28  
 Query Match: 79.25% Indels: 1  
 DB: 3 Gaps: 1

US-09-545-199F-4 (1-289) x AAC79654 (1-867)

Qy 1 MetAlaGlyAlaLysGluLeuThrLysLeuAlaSerValLysSerThrGlnLysIle 20  
 Db 1 ATGGCAGGTGCGAAGAGATAGAACCAAAATTCGAAGTGTGAANAATCTCAAAAATC 60  
 Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 Db 61 ACCAAGCAATGGAATGGTTGCTACTCTAAATCGGTAAACGCAAGAGCGTATGGCT 120  
 Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 121 GCCAGTCGCTTATTTCGGAACAATCCGTAGGTGATTAGCCATATTGCGAAGGAAGC 180  
 Qy 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
 Db 181 ATGGTTATAGCAACCCGTTTTTAACTGAACGCTGATATAAAAGTAGGCTACTCTGTC 240  
 Qy 81 ValSerThrAspArgGlyLeuGlyGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
 Db 241 GTTTCGACGATCGCGTTTATCGGTGCTTATATCAATTTATTCAAGCGACTTTG 300  
 Qy 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120  
 Db 301 AATGAATTTAAACGCTGGAAGATGAACGCTTAGTGTGCTGTTGTTAGTGGTCTG 360  
 Qy 121 LysSerIleAsnPheGlnSerLeuGlyLysLysLysLysLeuThrGlnAspSerGlyIle 140  
 Db 361 AAAGCGTAAGCTTTTACCAAAATCTAGCTTAAACGTGAGATCTCAAGTAACGGGATTA 420  
 Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160  
 Db 421 GCGGATAATCCGGAATGAACGATCTCGTGGCGCACTTAATGAATGATTATGCTTTC 480  
 Qy 161 LysLysGlyGluValAspValValLysValLysValLysPheLeuAsnThrMetSer 180  
 Db 481 CGAAACGGAGAGTGGATGGGTTTACGCTTACACCGCTTTTGAATAATACGATGTCA 540  
 Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
 Db 541 CAAAACCTGTTATCGCACAGTACTTCTCGTACCTTAACCTAGATGACGATGAATA--- 597  
 Qy 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
 Db 598 GATACGAAAGGTTTCATGGGATTAATTTATGAACCGAATCCCAAGTTTATTGGATAGT 657  
 Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240  
 Db 658 TTACTTGTTCGTTATTAGAACTCAGGTATACCAAGCAGTTGTAGTACCTAGCTTCT 717  
 Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260  
 Db 718 GAACAAGCCGCTCGAATGGTAGCGATGAAGCGCAACAGATATATCGGGTACATTAATC 777  
 Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 778 GATGAATTACAATTAGTGATTAACAAGCTCCCAAGCAAGCAATTAACAATTAATTAAC 837  
 Qy 281 GluIleValAlaGlyAlaAlaIle 289

Db 838 GAAATTGTTGCGGTGCGCGAGCAATT 864

RESULT 6  
 ABQ83531  
 ID ABQ83531 standard; DNA; 867 BP.  
 AC ABQ83531;  
 XX  
 DT 24-JAN-2003 (first entry)  
 XX  
 DE Actinobacillus pleuropneumoniae atpG gene SEQ ID NO:132.  
 XX  
 KW Antibacterial; vaccine; gram negative bacterial virulence gene;  
 KW identification; virulence; Pasteurellaceae; gene; ds.  
 OS  
 XX Actinobacillus pleuropneumoniae.  
 XX  
 PN WC200275507-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 17-JAN-2002; 2002WC-US001971.  
 XX  
 PR 15-MAR-2001; 2001US-00809665.  
 XX  
 PA (PHAA ) PHARMACIA & UPOHNS CO.  
 XX  
 PI Lowery DE, Fuller TE, Kennedy MJ;  
 XX  
 DR WPI; 2002-740868/80.  
 XX  
 PT P-PSDB; ABP54531.  
 XX  
 PS New mutant gram-negative bacteria, useful as vaccines and for identifying  
 XX new anti-bacterial agents that target virulence genes and their products.  
 CC Claim 5; Page 297-298; 350pp; English.

The present invention describes a gram-negative bacteria comprising a mutation in a gene, where the mutation results in decreased activity of a gene product encoded by the mutated gene. Also described is a method for producing a gram-negative bacteria mutant or an attenuated Pasteurellaceae bacteria. The mutated genes have antibacterial activity and can be used in vaccines. The gram-negative bacteria or the attenuated Pasteurellaceae bacteria can be used as vaccines in the fields of human medicine or veterinary medicine, and for identifying new antibacterial agents that target the virulence genes and their products. ABQ83458 to ABQ83578 and ABP54473 to ABP54551 represents sequences used in the exemplification of the present invention

Sequence 867 BP; 292 A; 147 C; 197 G; 231 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 7,71e-113 Length: 867  
 Score: 1132.50 Matches: 215  
 Percent Similarity: 89.97% Conservativity: 45  
 Best Local Similarity: 74.39% Mismatches: 28  
 Query Match: 79.25% Indels: 1  
 DB: 3 Gaps: 1

US-09-545-199F-4 (1-289) x ABQ83531 (1-867)

Qy 1 MetAlaGlyAlaLysGluLeuThrLysIleAlaSerValLysSerThrGlnLysIle 20  
 Db 1 ATGGCAGGTGCGAAGAGATAGAACCAAAATTCGAAGTGTGAANAATCTCAAAAATC 60  
 Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 Db 61 ACCAAGCAATGGAATGGTTGCTACTCTAAATCGGTAAACGCAAGAGCGTATGGCT 120  
 Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 121 GCCAGTCGCTTATTTCGGAACAATCCGTAGGTGATTAGCCATATTGCGAAGGAAGC 180

Fri Nov 12 17:22:00 2004

us-09-545-199f-4.p2n.rng

QY 61 IleGlyTyrLysHisPropheLeuValAspArgGluValLysValGlyMetIleVal 80  
 Db 181 ATTGGTTATAGACCCCGTTTAACTGAACGATGATTAATAAAGTGGCTATCTTGTTC 240  
 QY 81 ValSerThrAspArgGlyLeuGlyGlyLeuValAsnValAsnLeuPheLysThrValLeu 100  
 Db 241 GTTTCGACCGATCGCGTTTATGCGGTGCTTAATATCAATTTATCAAGCGACTTGG 300  
 QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGluLeuSerLeuIleGlySer 120  
 Db 301 AATGAATTTAAACCGTGAAGATAAAGCGTTAGTGTGAGCTTGGTTAGTGGTTCG 360  
 QY 121 LysSerIleAsnPhePheGlnSerLeuGlyLysIleLeuThrGlnAspSerGlyIle 140  
 Db 361 AAAGCGTAAGCTTTTACCAAAATCTAGCTTAACTGAGACTCAAGTAACGGGATTA 420  
 QY 141 GlyAspThrProSerValGluGluLeuLeuGlySerValAsnSerMetIleAspAlaTyr 160  
 Db 421 GCGGATAATCCGAAATCGAAGCTATCGTGGCGCAGTTAATGAATGATTAATGCGTTC 480  
 QY 161 LysLysGlyGluValAspValValLysValLysValLysPheIleAsnThrMetSer 180  
 Db 481 CGAAACGGGAAGTGGATCGCGTTTACGCTTACACCGCTTTTGAATATCAAGTGTCA 540  
 QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnGluLeuGly 200  
 Db 541 CAAAACCTGTTATCGCACAGTTACTTCGTTTACCTAACTAGATGACGATGAATTA 597  
 QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
 Db 598 GATACGAAAGGTTCTGCGGATTAATTTATGACCGAATCCACAGTTTATTTGGATAGT 657  
 QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240  
 Db 658 TTACTTGTTCGTTATTTAGAACTCAGGTATACCAAGCAGTTGTAGATAACCTAGCTTCT 717  
 QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260  
 Db 718 GAACAAAGCCGCTCGAATGGTAGCGATGAAGCCGCAACAGATAATGCGGGTACATTAATC 777  
 QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 778 GATGAATTAATAGTGTATACAAAGCTCGCAAGCAGCAATTAACAATGAATTAAC 837  
 QY 281 GluLeuValAlaGlyAlaAlaLalle 289  
 Db 838 GAAATGTTGCGGGTCCGCGACCAATT 864  
 RESULT 7  
 ABQ83549  
 ID ABQ83549 standard; DNA; 866 BP.  
 XX AC ABQ83549;  
 XX AC  
 DT 29-AUG-2003 (revised)  
 DT 24-JAN-2003 (first entry)  
 DE Pasteurella haemolytica atpG gene SEQ ID NO:166.  
 XX Antibacterial; vaccine; gram negative bacterial virulence gene;  
 XX identification; virulence; Pasteurellaceae; gene; db.  
 XX Mannheimia haemolytica.  
 XX WO200275507-A2.  
 XX PD  
 XX 26-SEP-2002.  
 XX PF 17-JAN-2002; 2002WO-US001971.  
 XX PR 15-MAR-2001; 2001US-00809665.  
 XX PHAA ) PHARMACIA & UPJOHN CO.

XX Lowery DE, Fuller TE, Kennedy MJ;  
 XX WPI; 2002-740868/80.  
 XX P-PSDB; ABP54547.  
 XX New mutant gram-negative bacteria, useful as vaccines and for identifying  
 XX new anti-bacterial agents that target virulence genes and their products.  
 XX Claim 5; Page 329-330; 350pp; English.  
 XX The present invention describes a gram-negative bacteria comprising a  
 XX mutation in a gene, where the mutation results in decreased activity of a  
 XX gene product encoded by the mutated gene. Also described is a method for  
 XX producing a gram-negative bacteria mutant or an attenuated  
 XX Pasteurellaceae bacteria. The mutated genes have antibacterial activity  
 XX and can be used in vaccines. The gram-negative bacteria or the attenuated  
 XX Pasteurellaceae bacteria can be used as vaccines in the fields of human  
 XX medicine or veterinary medicine, and for identifying new antibacterial  
 XX agents that target the virulence genes and their products. ABQ83458 to  
 XX ABQ83578 and ABP54473 to ABP54551 represents sequences used in the  
 XX exemplification of the present invention. (Updated on 29-AUG-2003 to  
 XX standardise OS field)  
 XX SQ Sequence 866 BP; 283 A; 150 C; 195 G; 238 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 6.92e-107 Length: 866  
 Score: 1077.50 Matches: 208  
 Percent Similarity: 87.89% Conservative: 46  
 Best Local Similarity: 71.97% Mismatches: 34  
 Query Match: 75.40% Indels: 1  
 DB: Gaps: 1  
 US-09-545-199F-4 (1-289) x ABQ83549 (1-866)  
 QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
 Db 1 ATGCGAGGTGTAAGAGATAAGAACCAAAATTCGTAAGTTCGTAATACACAAAAAT 60  
 QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnLysArgMetSer 40  
 Db 61 ACCAAAGCGATGGAATGTTGTCGCGCATCAAAATGCGTAAACCCCAAGCGGTATGGCG 120  
 QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 121 GCTTCTCGCCTTATGCTGAAAGTATTGCGAAGCAATTAGCCATATTGCCAAGGTAAAC 180  
 QY 61 IleGlyTyrLysHisPropheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
 Db 181 ATTGAGTATTAACACCCCAATTTTACCCCGCTCCGTTAAAGAAAGTTGCTATTAGTA 240  
 QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
 Db 241 GTTTCACCGCATCGCGTTTATGTTGGCTTAATATCAATTTATTTAAACCGTTT 300  
 QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120  
 Db 301 CATGAATGAAAGAAAAGATGACCAAGTGTTAAGTCTCGACTTCGCTGGTGGGAAT 360  
 QY 121 LysSerIleAsnPhePheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140  
 Db 361 AAAGGATCTCTTTTAAACCAATGGGCTAGAGATTAAGATTAAGGTTCATATCAATGGATT 420  
 QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160  
 Db 421 GGTGATACACCGCAATGGAAGATTAGTCGGTATTGTTAATGTTGTTAATGTTAATGTTTCA 480  
 QY 161 LysLysGlyGluValAspValValTyrLysValTyrAsnLysPheIleAsnThrMetSer 180  
 Db 481 CGTGAAGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnGluLeuGly 200

```

Db 541 CAAAACCCGACAGTCAACAGTTGCTTCTTGGCTGCTGCACTGGAAATGACTCATTA--- 597
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 598 GAGCAAACTGGTTCTTGGGATTTATCTCATGAACCAATCCACAAGCGTTATTAGACAGC 657
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaLaValGluAsnLeuAlaSer 240
Db 658 TTTACTGGTTCGTATTAGATCTCAAGTTATCAGGCAGTGGTAGATAATCTTGGCTCT 717
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLe 260
Db 718 GAACAGGCTGCTCGAATGCTGCAATGAATGAACGACGCAACCGATAACGCGAGGTATCTGATT 777
QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db 778 AATGAGTTACAGTTAGTGTATACCAAGCTCGTCAAGCAAGTATACGAATGAATTAAT 837
QY 281 GluLeuValAlaGlyAlaAlaLeile 289
Db 838 GAAATTTGCGGGTGGCGCGCAAT 864

RESULT 8
ACF70377 standard; DNA; 864 BP.
XX AC ACF70377;
XX XX
XX 20-NOV-2003 (first entry)
XX DE Photorhabdus luminescens nucleotide sequence #8844.
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX KW detection; food; gene expression; plant; animal; microorganism; toxin;
XX KW antibiotic; biopesticide; virulence factor; disease model; plague;
XX KW whooping cough; gene; ds.
XX OS Photorhabdus luminescens.
XX XX
XX PN WO200294867-A2.
XX XX
XX 28-NOV-2002.
XX PF 07-FEB-2002; 2002WO-IB003040.
XX XX
XX 07-FEB-2001; 2001FR-00001659.
XX PR (INSP ) INST PASTEUR.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX PI
XX PA Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX PI Buchrieser C;
XX XX
XX WI; 2003-148459/14.
XX DR
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX XX
XX PS Claim 2; SEQ ID NO 8844; 1205pp; French.
XX XX
XX CC The invention relates to the isolation of genes and their encoded
XX CC proteins from Photorhabdus luminescens. The isolated sequences are
XX CC sources of probes and primers for detecting the genome of P. luminescens
XX CC and related species; to study polymorphisms; for gene analysis and for
XX CC detection/amplification of the genes. Antibodies (Ab) raised against the
XX CC polypeptides encoded by the genes are used for detection/identification
XX CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX CC carry a gene-containing vector are used to select compounds that
XX CC modulate, regulate, induce or inhibit expression of the genes in plants,
XX CC animals or microorganisms other than P. luminescens and are able to alter
XX CC response or sensitivity to toxins and antibiotics produced by P.
XX CC luminescens. Cells transformed to express the genes are useful for

```

```

CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX SQ Sequence 864 BP; 236 A; 183 C; 229 G; 216 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.25e-99 Length: 864
Score: 1004.00 Matches: 195
Percent Similarity: 84.08% Conservative: 48
Best Local Similarity: 67.47% Mismatches: 44
Query Match: 70.26% Indels: 2
DB: 10 Gaps: 1

US-09-545-199F-4 (1-289) x ACF70377 (1-864)
QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
Db 1 ATGGCGCGCGCAAAAGAGATACGTACCAAGATGCCAGCGTGCAGAAACACGCAAAAATC 60
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
Db 61 ACTAAAGCGATGGAGATGGTTGCTGCGTCCAAAATGCGTAAACGCGAGGATCCATGGCG 120
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db 121 GCCACGCGTCTTATGCGAGAAACCATACGCGCGTGATTGCTCACCCTGGTAGGTAAT 180
QY 61 IleGlyTyrLysHisProPheLeuValAspArgLysValLysValGlyMetIleVal 80
Db 181 CTGGAATATAAACATCCGTACCTTGAAGAGCGGTGAAACCAACGCTGCGGGTATCTGGTT 240
QY 81 ValSerThrAspArgGlyLeuCysGlyLeuAsnValAsnValAsnLeuPheLysThrValLeu 100
Db 241 GTTTCACCGATCGTGGCTTGTGCGGTGTTTGAATACTAATCTGTTTCAAAAACCTGCTT 300
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db 301 TCAGAAATGAAGACTGGTCTGATAAAGATGCCAGTGTGAGTGGCGCTATTCGGCTCT 360
QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db 361 AAGGCTACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 420
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 421 GGAGATAACCCCTTCACTCTCTGAATTCGATCGCGGCGAGTGAATATCATCTGCGGGCATAC 480
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 481 GATGAGGAGCGTCTGGATAAACTGTATGTGTGTGACAAACAGTTCATCATCATCATGTC 540
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 541 CAGGAACCGACTATTACTCAGTTATTACTCTGCTGCGGAGATGATGAACACACTG--- 597
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 598 ---AGAGAAATCTCTGGGATTTATTTACCAAGCTGATCTCAAGCGGTGCTAGATATA 654
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaLaValGluAsnLeuAlaSer 240
Db 655 CTGCTGGTTCGTATTAGTGAATTCGAGGTTTATCAGGCGCGTGGTTGAAACCTGGCTAGT 714
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLe 260
Db 715 GAACAGGCGCGCAAGATGGTGGCGGATGAAGCCCAACTGATATATGTTGGCAGCCTGATC

```



Score:	1004.00	Matches:	195
Percent Similarity:	84.08%	Conservative:	48
Best Local Similarity:	67.47%	Mismatches:	44
Query Match:	70.26%	Indels:	2
DB:	10	Gaps:	1
US-09-545-199F-4 (1-289) x ACF65385_5 (1-110000)			
QY	1	MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle	20
DB	60955	ATGGCGCGCGCAAAAGAGATACGTACCAAGATCGCCAGCGTGCAAAACGCAAAAAATC	61014
QY	21	ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer	40
DB	61015	ACTAAAGCGATGGAGATGGTCTCGCTCCAAAATGCGTAAACGCGAGGATCGCATGGCG	61074
QY	41	SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr	60
DB	61075	GCACGCGCTCTTATGCGAAGAACCATACGCGCGTGTGCTCACCTTGGCTTAGGTAAAT	61134
QY	61	IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal	80
DB	61135	CTGGAATATAAATCCGTACCTTGAAGAGCGTGAACCAACGATGTCGGGTATCTGGTT	61194
QY	81	ValSerThrAspArgGlyLeuGlyGlyLeuAsnValAsnLysPheLysThrValLeu	100
DB	61195	GTTCCTACCGATCGTGGCTGTGCGGTGGTTGATATCTATCTGTCAAAAACCTGCTT	61254
QY	101	AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer	120
DB	61255	TCAGAAATGAAGACTGGTCTCTGATAAAGATGTCAGTGTGAGCTGGCGCTTATCGGCTCT	61314
QY	121	LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle	140
DB	61315	AAGGCTACCTCTTCTTCCTCTCTGTGGGGTAACGTGTGCTCAGGTAAACAGGCATG	61374
QY	141	GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr	160
DB	61375	GGAGATAACCTTCACTGCTGAATGTATCGGCGCAGTGAATATCATGCTGCGGCGATAC	61434
QY	161	LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer	180
DB	61435	CATGAGGACCTCGGATTAATATGATGTTGTTGTCACAAACAGTTCATCAATCAATGTC	61494
QY	181	GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnGluLeuGly	200
DB	61495	CAGAAACCGACTATTACTCAGTTATTAATCTCTGCTGCGGAGATGATGAACACTG---	61551
QY	201	GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn	220
DB	61552	---AAGAGAAATCCTGGGATATTATTAACGAACTGATCTTAAGCGCTTGTAGATATA	61608
QY	221	LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer	240
DB	61609	CTGCTGCGCTGTTATGAAATCGCAGGTTTATCAGGGCGCTGTTGAAACCTCGGCTAGT	61668
QY	241	GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle	260
DB	61669	GAACAGCGCGACGAATGGTGGCGATGAAGCGCAACTGATATATGGTGGCAGCTGATC	61728
QY	261	AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn	280
DB	61729	AAAGAGTTGCGAGTTGGTTTATTAACAAAGCTCGTCAAGCCAGCATAACTCAGGAGCTGACC	61788
QY	281	GluIleValAlaGlyAlaAlaIle 289	
DB	61789	GAGATTGCTCGGGTGTCTCCGCGGTT 61815	
RESULT 11			
ACF67367_34			
Continuation (35 of 57) of ACF67367 from base 3400001 (Photorhabdus luminescens nucleoti			
WP	Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367		
WP	Fragment Name Begin End		

```
Db 66149 ACTAAAGCGATGAGATGGTGTGCTGCTCAAAATCGTAAACACGAGATCGCATGGCG 66208
Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db 66209 GCCAGCGCTCTTATGCAGAACCATACGACGCGTGTGGTCACTTGGCTAGGTAAAT 66268
Qy 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal 80
Db 66269 CTGGAATATAAACATCCGACCTTGAAGAGCGTGAACCAACGTCGTGGGTATCGGTT 66328
Qy 81 ValSerThrAspArgGlyLeuCysGlyValLeuAsnValAsnLeuPheLysThrValLeu 100
Db 66329 GTTCTACCGATCGTGGCTGTGGCTGTGGTGTGAATCTAATCTGTTCAAAACGCTGT 66388
Qy 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db 66389 TCAGAAATGAAGACCTGGTCTGATAAAGATGCCAGTGTGAGCTGGCGCTTATCGGCT 66448
Qy 121 LysSerIleAsnPhePheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db 66449 AAGCTACCTCTTCTTCTGCTCTGTTGGGTAAGCTGTGTGCTCAGTAAACGGCATG 66508
Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 66509 GGAGATAACCTTCACTGCTGAATGTATGTCGGCCAGTGAATATCATGCTCGGGCATC 66568
Qy 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 66569 GATGAGGACGCTGTGATAACTGTATGTGTGACAAACAGTTCAATACATGTCC 66628
Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnGluLeuGly 200
Db 66629 CAGGAACCGCATTTACTCAGTTATTACTCTGCTGCGCGAGATGATGAACACTG-- 66685
Qy 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAsn 220
Db 66686 ---AAGAAATCCCTGGGATTTATTATACGAACTTCCTAAGCGGTGCTAGATATA 66742
Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240
Db 66743 CTGCTCGCTCGTTATGATAGATCGCAGGTTTATCAGGCGCTGTTGAACCTGGCTAGT 66802
Qy 241 GluGlnAlaAlaArgMetValAlaValLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db 66803 GAACAGCCGCGACGAATGTGGCGATGAAGCCGCACTGATAATGGTGCAGCGCTGATC 66862
Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
Db 66863 AAAGAGTTGAGTTGTTTATAACAAAGCTGCTCAGGCCAGCATAACTCAGGAGCTGACC 66922
Qy 281 GluIleValAlaGlyAlaAlaIle 289
Db 66923 GAGATTGTCTCGGGTCTCCGCGTT 66949
RESULT 12
AAH81345
ID AAH81345 standard; DNA; 864 BP.
XX
AC AAH81345;
XX
DT 21-SEP-2001 (first entry)
XX
DE Escherichia coli protein encoding nucleotide sequence SEQ ID NO:144.
XX
KW Escherichia coli; identification; proliferation; gene therapy; diagnosis;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW bacterial growth inhibition; ds.
XX
OS Escherichia coli.
XX
PN WC0200148209-A2.
XX
```

```
PD 05-JUL-2001.
XX
XX 19-DEC-2000; 2000WO-US034419.
XX
XX 23-DEC-1999; 99US-017300SP.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Forsyth RA, Ohlsen KL, Zyskind JW;
XX
XX WPI: 2001-457376/49.
XX
XX P-PSDB; AAG98289.
XX
XX Novel nucleic acids encoding proteins required for Escherichia coli
XX proliferation, useful for screening for antimicrobial agents.
XX
XX Claim 9; Page 191-192; 596pp; English.
XX
XX The present invention describes a purified or isolated nucleic acid
XX sequence (I) consisting essentially of one of the 93 nucleotide sequences
XX given in AAH81202 to AAH81294, where expression of the nucleic acid in a
XX microorganism is capable of inhibiting proliferation of a microorganism.
XX (I) have antibacterial and antibiotic activities, and can be used in gene
XX therapy. Expression of (I) in a microorganism inhibits proliferation of
XX the microorganism, and the manufactured antibiotic is useful for reducing
XX the activity or level of a gene product required for proliferation of a
XX microorganism in a subject, specifically humans. The nucleic acids that
XX inhibit bacterial growth or proliferation can be used as antitense
XX applications, the nucleic acid sequences complementary to sequences
XX required for proliferation can be used as diagnostic tools. For example,
XX nucleic acid probes complementary to proliferation-required sequences
XX that are specific for particular species of microorganisms can be used as
XX probes to identify particular microorganism species in clinical
XX specimens. AAH81295 to AAH81487 encode the Escherichia coli proteins
XX given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent
XX oligonucleotides, which are used in the exemplification of the present
XX invention
XX
XX SQ Sequence 864 BP; 220 A; 222 C; 240 G; 182 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1.6e-97 Length: 864
XX Score: 991.00 Matches: 191
XX Percent Similarity: 82.35% Conservative: 47
XX Best Local Similarity: 66.09% Mismatches: 49
XX Query Match: 69.35% Indels: 2
XX DB: 5 Gaps: 1
XX
XX US-09-545-199F-4 (1-289) x AAH81345 (1-864)
XX
XX Qy 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
XX Db 1 ATGGCCGCGCAAAAGAGATACGTAGTACGAGCGTCCAGAACACGCAACGCAAGATC 60
XX
XX Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
XX Db 61 ACTAAAGCGATGAGATGGTGGCGCTTCCAAATCGTAAATCGCAGATCGCATGGCG 120
XX
XX Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
XX Db 121 GCCAGCGCTCTTATGCAGAACCATCGCAGATGATGGTCACTTGCACACCGTAAAT 180
XX
XX Qy 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal 80
XX Db 181 CTGGAATATAAACATCCGACCTTACCTTGAAGACCGCGAGTAAACGCGTGGGTACCTGGT 240
XX
XX Qy 81 ValSerThrAspArgGlyLeuCysGlyValLeuAsnValAsnLeuPheLysThrValLeu 100
XX Db 241 GTGTGACCGACCGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 300
XX
XX Qy 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
XX Db 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
XX
```





Db 805 AAGAGCTCCAGTTGGTATACAAAGCTCGTCAGGCGAGCATTAATCGGAAGTCAAC 864  
 Qy 281 GluileValAlaGlyAlaAlaAlaAla 289  
 Db 865 GAGATCGTCTCGGGGGCGCGCGGT 891

## RESULT 14

ABD03905  
 ID ABD03905 standard; DNA; 876 BP.

XX ABD03905;  
 AC ABD03905;  
 DT 29-JUL-2004 (first entry)  
 XX Pseudomonas aeruginosa polynucleotide #2509.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 XX antibacterial.

OS Pseudomonas aeruginosa.

PN US6551795-B1.

PD 22-APR-2003.

PF 18-FEB-1999; 99US-00252991.

PR 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

DR WPI; 2003-615309/58.

DR P-PSDB; ABO70334.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.

PS Disclosure; SEQ ID NO 2509; 455pp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABD01397.  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from the printed  
 CC seqdata.uspro.gov/sequence.html

SQ Sequence 876 BP; 206 A; 250 C; 266 G; 154 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1-56e-90 Length: 876  
 Score: 926.50 Matches: 180  
 Percent Similarity: 79.58% Conservative: 50  
 Best Local Similarity: 62.28% Mismatches: 56  
 Query Match: 64.84% Indels: 3  
 DB: 11 Gaps: 1

US-09-545-199F-4 (1-289) x ABD03905 (1-876)

Qy 1 MetAlaGlyAlaLysGluileArgThrLysileAlaSerVallySerThrGlnLysile 20  
 Db 16 ATGGCAGCGCCAAAGAGATTGCGCAGGATTGCGAGCATCAAAAGCAGCGAAAAGATC 75  
 Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 Db 76 ACCAATGCCATGGAAAAAGTGGCGGTGAGCAAGATGCGCAGGCAAAATGCGCATGGCG 135  
 Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 136 GCCGCCGTCCCTACCGGAGCGTATTTCGCCAGGTGATCGCCATCTGCCAACGCGCAAC 195  
 Qy 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
 Db 196 CCGGAATACCGTCACCGCTTCATGGTCGAGCGTGAAGTCAAGCGGCTCGCTACATCGTG 255  
 Qy 81 ValSerThrAspArgGlyLeuCysGlyLysValAsnValAsnLeuPheLysThrValLeu 100  
 Db 256 GTGAGCAGTGACCGTGTCTGTCGCGCGGTCTGAACATCAACCTGTTCAAGTCCCTCGTC 315  
 Qy 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120  
 Db 316 AAGCAGATGAGCGCTACCGCAGCAAGGGCGGAATCGACCTTTCGCTTCGTATCGTAGC 375  
 Qy 121 LysSerIleAsnPhePheGlnSerLeuGlyIleLysileLeuThrGlnAspSerGlyIle 140  
 Db 376 AAGGCGCTTCGTTCTTCGCGAGCTTTCGCGCGCAACGTCGTCGCGACTATCAGCCACCTC 435  
 Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaThr 160  
 Db 436 GCGAAGAGCGCTTCGATCAACGACCTGATCGCGAGTGTCAAGGTGATGCTCAGCAGTAC 495  
 Qy 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
 Db 496 CTCGAAGGCGGTATCGATCGCTTTCGTTGCTCCAAACAAAGTTCGTCAACACCATGACC 555  
 Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
 Db 556 CAGAGCGCGACCGTGAACAGCTGATTCGCTGGTGGCGGATGACGATCAGGAGCTG--- 612  
 Qy 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
 Db 613 -----AAGCACCCTGGGACTATCTCTACGAACCGCAGCCCAAGTCTCTCTCGACGGG 666  
 Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240  
 Db 667 CTGCTGTCGCTTACGTGAATCCAGGTGTACCGAGCGCGTGGTGTGAGAACACCGCTGT 726  
 Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260  
 Db 727 GAGCAGCGCGCGCGATGATTGCAATGAAGAAGCTACCGCAACCGCGCGGAGCTGATC 786  
 Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 787 AGCGATTTCACCTGATCTACACAGCGCGGTGAGCGCGATCAGCAGGAAATCTCG 846  
 Qy 281 GluileValAlaGlyAlaAlaAlaIle 289  
 Db 847 GAAATCGTCGGCGCGCTCGCGCGTG 873

## RESULT 15

ABD03801/c

ID ABD03801 standard; DNA; 1050 BP.

XX ABD03801;

DT 29-JUL-2004 (first entry)

XX Pseudomonas aeruginosa polynucleotide #2405.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 XX antibacterial.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 9, 2004, 07:03:13, Search time 24316 Seconds  
(without alignments)  
433.093 Million cell updates/sec

Title: US-09-545-199F-4

Perfect score: 1429

Sequence: 1 MAGAKEIRTKASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2 1/USPTO.spool p/US09545199/runat 08112004 112526 20963/app\_query.fasta\_1.455  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09545199 @cgn 1 1 6425 @runat 08112004 112526 20963 -NCFU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	693	48.5	699	4	BJ498955
C 2	675	47.2	728	7	CF983748
C 3	605.5	42.4	1469	8	BZ572880
C 4	463	32.4	1437	3	AY108268
C 5	440	30.8	1313	3	CNS0A3T2
C 6	432.5	30.3	807	8	BH387684
C 7	406	28.4	337	8	BH317803
C 8	403	28.2	796	8	BH393596
C 9	381	26.7	1474	8	BZ569821

C 10	379.5	26.6	448	8	AQ936809
C 11	352.5	24.7	666	8	BH378165
C 12	350.5	24.5	1205	3	CNS0A3BS
C 13	349.5	24.5	667	8	BH385025
C 14	342.5	24.0	872	3	CF651710
C 15	340	23.8	1514	3	AY108441
C 16	337.5	23.6	852	7	CN820906
C 17	335	23.4	916	7	CO004140
C 18	335	23.4	917	7	CO008879
C 19	335	23.4	928	7	CO031253
C 20	335	23.4	987	7	CO001829
C 21	334.5	23.4	1064	6	CB686051
C 22	333.5	23.3	838	7	CO364454
C 23	333.5	23.1	358	4	BG876920
C 24	330	23.0	904	5	BU797110
C 25	329	23.0	1036	2	BF942618
C 26	326.5	22.8	858	7	CO088347
C 27	323	22.6	847	4	BG839438
C 28	319.5	22.4	822	7	CK788660
C 29	318.5	22.3	836	7	CO078673
C 30	318	22.3	869	5	EX742945
C 31	318	22.3	872	7	CK191546
C 32	318	22.3	886	7	CK189737
C 33	318	22.3	890	7	CK189738
C 34	318	22.3	898	7	CK189587
C 35	318	22.3	911	7	CK188511
C 36	316	22.1	835	7	CR449718
C 37	316	22.1	876	7	CR414264
C 38	314	22.0	877	7	CO082073
C 39	313	21.9	846	7	CR569311
C 40	313	21.9	1213	3	AY439378
C 41	312.5	21.9	1064	4	BG837916
C 42	312.5	21.9	1083	3	AY375344
C 43	310	21.7	1085	3	AK088164
C 44	310	21.7	1736	3	BC048777
C 45	309	21.6	1104	7	CN061585

#### ALIGNMENTS

RESULT 1  
BJ498955/c  
LOCUS  
DEFINITION BJ498955 MF01FSA cDNA Oryzias latipes cdna linear EST 08-AUG-2002  
mRNA sequence.  
VERSION BJ498955.1 GI:22150917  
KEYWORDS EST.  
SOURCE Oryzias latipes (Japanese medaka)  
ORGANISM Oryzias latipes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Telostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
REFERENCE 1 (bases 1 to 699)  
Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.  
Medaka EST Project in Takeda's lab  
Unpublished (2001)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1. 699

/organism="Oryzias latipes"

/mol\_type="mRNA"

/strain="d-rr"

/db\_xref="taxon:8090"

/clone="MF01FSA015K15"

/sex="mixture of female and male"

#### FEATURES

source

/tissue type="whole embryo"  
/dev stage="fry stage 40"  
/clone\_lib="XF01PSA CDNA"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,75e-73 Length: 699  
Score: 693.00 Matches: 134  
Percent Similarity: 79.33% Conservative: 31  
Best Local Similarity: 64.42% Mismatches: 41  
Query Match: 48.50% Indels: 2  
DB: 4 Gaps: 1

US-09-545-199F-4 (1-289) x BJ498955 (1-699)

QY 82 SerThrAspArgGlyLeuGlyGlyLeuAsnValAsnLeuPheLysThrValLeuAsn 101  
Db 698 TCAGCCAGCGTGTGTGCGGTGTGGAACATTAACTCTTCAAAAAGCTGCTGGCG 639  
QY 102 GLuMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuLeuSerLys 121  
Db 638 GAATGAGACCTGGACCGCAAGCGGTCAATGCGACCTCGCAATGCTCGAAA 579  
QY 122 SerLeuAsnPheGlnSerLeuGlyLysLeuThrGlnAspSerGlyLys 141  
Db 578 GCGGTGTGCTTCAACTCGGCGCGCAATGTGTGCCAGGTACCGGATGGG 519  
QY 142 AspThrProSerValGluGlnLeuLeuGlySerValAsnSerMetLeuAspAlaTrpLys 161  
Db 518 GATACCTTCCTGTCGCAAGATGATGCGTCCGTAAGTGTGTCAGGCTACGAC 459  
QY 162 LysGlyGluValAspValValTrpLeuValTrpAsnLysPheLeuAsnThrMetSerGln 181  
Db 458 GAAGCGCGTCTGACAAAGCTTTACATTGTACCAACAATTTATTAACCATGCTCTCAG 399  
QY 182 LysProValLeuGluLysLeuLeuProLeuProGluLeuAspAspGluLeuGlyGlu 201  
Db 398 GTTCGACCATCAGCCAGCTGTGCGGTGTACCGCATCAGATGATGATCTG----- 345  
QY 202 ArgLysGlnValTrpAspTrpLysLeuValProAspAlaLysValLeuLeuAspAsnLeu 221  
Db 344 AAACATAAATCTGGATTACCTGTACGACCGGATCCGAGGCGTGTGCTGATACCCCTG 285  
QY 222 LeuValArgTrpLeuGluSerGlnValTrpGlnAlaValGluLeuAlaSerGlu 241  
Db 284 CTGCTGCTGATATGCGAATCTCAGGTTTATCAGGCGGTGTGAAACCTGGCCAGCGAG 225  
QY 242 GlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLeuAsn 261  
Db 224 CAGGCGCGCGTATGTCGCGATGAAGCCGCGACCAATGCGCGACGCTGTGATATAA 165  
QY 262 GluLeuGlnLeuValTrpAsnLysAlaArgGlnAlaSerLeuThrAsnGluLeuAsnGlu 281  
Db 164 GAGCTGCGAGTGGTATACAAAGAGCTGTCAGGCCAGCATTAATCTCAGGAACCTCAGCGAG 105  
QY 282 IleValAlaGlyAlaAlaAlaIle 289  
Db 104 ATCGTCTCGGCGCGCGCGCGGT 81

## RESULT 2

CF983748/c

LOCUS

DEFINITION

CF983748

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 729)

CF983748 729 bp mRNA linear EST 25-NOV-2003  
ID05 grape selected express library Vitis vinifera CDNA 5', mRNA  
sequence.

CF983748.1 GI:38516573

EST

Vitis vinifera

Vitis vinifera

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; Vitaceae; Vitis.

1 (bases 1 to 729)

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Ablett, E., Seaton, G., Scott, K., Shelton, D., Graham, M.W.,  
Bavestock, P., Lee, L.S. and Henry, R.  
Analysis of grape ESTs: global gene expression patterns in leaf and  
berry  
Plant Sci. 159 (1), 87-95 (2000)  
Contact: Patricio Arce-Johnson  
Laboratory of Biochemistry, Department of Molecular Genetics and  
Faculty of Biological Sciences, Catholic University of Chile  
Alameda 340, Santiago, Chile  
Tel: (56-2) 6862897  
Fax: (56-2) 2225515  
Email: parce@genes.bio.puc.cl  
Seq primer: T3  
Location/Qualifiers  
1. 729  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/cultivar="Chardonnay"  
/db\_xref="taxon:29760"  
/clone\_lib="grape selected express library"  
/note="Organ: leaf and berry; Vector: pBluescript II SK+;  
Site 1: 5' Eco R I; Site 2: 3' Xho I; From leaves total  
RNA were isolated as described in Plant Mol Biol Reporter  
11 (1993) 212-215. From berries total RNA were isolated  
using CTAB method and lithium chloride precipitation  
(Plant Mol Biol Reporter 11 (1993) 117-121). Double  
stranded cDNA were prepared from 5 ug of polyA leaf RNA or  
25 ug total RNA berry RNA using a Stratagene cDNA  
synthesis kit. This method produces hemimethylated cDNAs  
with 5' EcoRI and 3' XhoI restriction sites. These cDNA were  
cloning in pBluescript SK II+. Following ligation,  
plasmids were transformed into XL1-blue-MRF' cells  
(Stratagene) and plated for blue-white selection."

## FEATURES

## source

## ORIGIN

## Alignment Scores:

Pred. No.: 4,47e-71 Length: 729  
Score: 675.00 Matches: 140  
Percent Similarity: 78.38% Conservative: 34  
Best Local Similarity: 63.06% Mismatches: 42  
Query Match: 47.24% Indels: 6  
DB: 7 Gaps: 1

US-09-545-199F-4 (1-289) x CF983748 (1-729)

QY 72 GluValLysLysValGlyMetIleValVal-SerThrAsp-ArgGlyLeuGlyGlyL 91  
Db 727 GACGTTAAACGCGTGGGCTACCTGGTGGTGTGCGACCGCCGCTGTGTGCGGGT 668  
QY 91 euAsnValAsnLeuPhe-LysThrValLeuAsnGluMetLysGluTrpLysGlyLysAsp 110  
Db 667 TGAACATTAACTGTTCATAAAACCTGCTGGCGGAAATGAAGACCTGGACCAAGGC 608  
QY 111 ValSerValGlnLeuSer-LeuIleGlySerLysSerIleAsnPheGlnSerLeuG 130  
Db 607 GTTCAATGCGACCTCGAAATGATCGCTCGAAAGCGGTGCTCTTCAACTCGGTGG 548  
QY 130 YIleLysIleLeuThrGlnAspSerGlyIleGlyAspThrProSerValGluGlnLeu 150  
Db 547 CGCAATGTGTGTGCCAGNTCACCGGATGGGGATAACCTTCCCTGTCGGAAGT 488  
QY 150 eGlySerValAsnSerMetIleAspAlaTrpLysLysGlyGluValAspValValTr 170  
Db 487 CGTCCGGTAAAGTATGTTTCAGCCCTACGACGAGCGCTGTGGACAACTTTACAT 428  
QY 170 uValTrpAsnLysPheIleAsnThrMetSerGlnLysProValLeuGluLysLeuLe 190  
Db 427 TGCACCAACAATTTATTAACACCATCTCTCAGGTTCCGACCATCAGCGCTGCTGCC 368  
QY 190 oLeuProGluLeuAspAsnAspGluLeuGlyGluArgLysGlnValTrpAspTrpL 210  
Db 367 GTTACCGGATCAGATGATGATCTG-----AAACATAAATCTCGGATACCTGTA 314

QY 210 rGluProAspAlaLysValLeuLeuAspAsnLeuValArgTyrLeuGluSerGlnVa 230  
 Db 313 CGAACCGATCCGAAGCGTGTGGATACCTCGCTGCTGTATGTGATCTCAGGT 254  
 QY 230 lTyrGlnAlaAlaValGluAsnLeuAlaSerGluGlnAlaAlaArgMetValAlaMetly 250  
 Db 253 TTATCAGGGCGTGTGTAACCACTGGCCAGCAGCAGCGCCCGTATGTCGCGATGAA 194  
 QY 250 sAlaAlaThrAspAsnAlaGlyAsnLeuLeuLeuLeuLeuLeuLeuValTyrAsnLysAl 270  
 Db 193 AGCCGACCGACCAATGCGCGACCTGTATTAAAGAGCTGCGATTTGGTATACACAAAGC 134  
 QY 270 aArgGlnAlaSerIleThrAsnGluLeuAsnGluLeuValAlaGlyAlaAlaLalle 289  
 Db 133 TCGTCAGGCAGCATTAATCAGGAACCTACCGAGATCGTCTCGGGGCGCGCGGTT 76

RESULT 3  
 BZ572880 1469 bp DNA linear GSS 17-DEC-2002  
 LOCUS  
 DEFINITION msh2\_2842.x3 msh Pseudomonas aeruginosa genomic clone msh2\_2842, genomic survey sequence.  
 ACCESSION BZ572880  
 VERSION BZ572880.1 GI:27207941  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 1469)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence Variation among multiple isolates of Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol. (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun

## FEATURES

source

1..1469  
 /location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone\_lib="msh2\_2842"  
 /note="Environmental isolate. Whole genomic shotgun library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3 33e-62 Length: 1469  
 Score: 605.50 Matches: 120  
 Percent Similarity: 72.57% Conservative: 44  
 Best Local Similarity: 53.10% Mismatches: 58  
 Query Match: 42.37% Indels: 4  
 DB: 8 Gaps: 1

US-09-545-199f-4 (1-289) x BZ572880 (1-1469)

QY 15 LysSerThrGlnLysIleThrLysAlaMetGluMetValAlaAlaSerLysMetArgLys 34  
 Db 43 AGGTCGACGGTATCGATAACCTTGATATCGAATTCCTGCAGCCGACGAAGATGGCAAG 102  
 QY 35 ThrGlnGluArgMetSerSerArgProTyrSerGluThrIleArgAsnValIleSer 54  
 Db 103 GCACAAATGCGATGGCGCGCGCGCTCCCTACGCGAGCGTATTCGCGAGTATCGCG 162  
 QY 55 HisValSerLysAlaThrIleGlyTyrLysHisProPheLeuValAspArgGluValLys 74

Db 163 CATCTGGCCAAACCCCAACCCGGAATACCGTACCGCTTCATGTCGAGCTGAAGTCAAG 222  
 QY 75 LysValGlyMetIleValValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsn 94  
 Db 223 CGCGTCGGCTACATCGTGTGTGAGCAGTACCGTGTCTGTGCGCGGCTCTGAACATCAAC 282  
 QY 95 LeuPheIysThrValLeuAsnGluMetLysGluIrrLysGluLysAspValSerValGln 114  
 Db 283 CTCTTCAAGTCCTCGTCAAGGACATGAGCGCTACCGCGAGCAAGGGCGGAAATCGAC 342  
 QY 115 LeuSerLeuIleGlySerLysSerIleAsnPheGlnSerLeuGlyIleLysIleLeu 134  
 Db 343 CTTTCGGTGTATCGTAGCAAGGCGCTTCGTCTTCCCGAGTTTCGGCGGCAACGTCGTG 402  
 QY 135 ThrGlnAspSerGlyIleGlyAspThrProSerValGluGlnLeuIleGlySerValAsn 154  
 Db 403 GCAGCTATCAGCACCTCGCGGAGAGAGCTTCGATCAACGACCTGATCGGCAGTGTCAAG 462  
 QY 155 SerMetIleAspAlaTyrLysIleGlyGluValAspValValTyrLeuValTyrAsnLys 174  
 Db 463 GTGATGCTCGACGCATACCTCGAAGCGCTATCGATCGCTGTTCGGTGTCTCCAAACAG 522  
 QY 175 PheIleAsnThrMetSerGlnLysProValLeuGluLysLeuIleProLeuProGluLeu 194  
 Db 523 TTTCGTCACACCATGACCCAGAGCCGACCGCTGGAACAGCTGATTCGCTGTCGCGCAT 582  
 QY 195 AspAsnAspGluLeuGlyGluArgLysGlnValTyrAspTyrIleTyrGluProAspAla 214  
 Db 583 GACGATCAGGAGCTG-----AAGCACCATCGGAGCTATCTCTACGAACCGCAGCC 633  
 QY 215 LysValLeuLeuAspAsnLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAl 234  
 Db 634 AAGTCGCTCTCGACGGGCTGCTGTCGGTACGTGGAATCCCAAGTGGACCAAGGCGCGG 693  
 QY 234 aValGluAsnLeuAla 239  
 Db 694 GGTGAGGACCAACGCC 709

## RESULT 4

AY108268

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
 If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at WSL, maizenap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.  
 Location/Qualifiers  
 1..1437  
 /organism="Zea mays"  
 /mol\_type="mRNA"

## FEATURES

source

1437 bp mRNA linear HTC 16-OCT-2002  
 Zea mays PC0078526 mRNA sequence.  
 AY108268  
 AY108268.1 GI:21211346  
 HTC.  
 Zea mays  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 1437)  
 Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.  
 Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes  
 Unpublished (2002)  
 2 (bases 1 to 1437)  
 Coe,E.H.  
 Direct Submission  
 Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
 If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at WSL, maizenap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.  
 Location/Qualifiers  
 1..1437  
 /organism="Zea mays"  
 /mol\_type="mRNA"

```

/db_xref="MaizeDB:634352"
/db_xref="taxon:4577"
/Clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

ORIGIN
Alignment Scores:
Pred. No.: 6,39e-45 Length: 1437
Score: 463.00 Matches: 114
Percent Similarity: 55.49% Conservative: 63
Best Local Similarity: 35.74% Mismatches: 108
Query Match: 32.40% Indels: 34
DB: 3 Gaps: 5

US-09-545-199F-4 (1-289) x AY108268 (1-1437)
QY 5 LysGluLeuArgThrLysLeuAlaSerValLysSerThrGlnLysIleThrLysAlaMet 24
DB 213 CGTGAGCTCGACCCGCGATCGCTCGGTCAAGAACGCGAGATCAGCGGCGCATG 272
QY 25 GluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSerSerArgPro 44
DB 273 AAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 332
QY 45 TyrSerGluThrIleArgAsnValIleSerHisValSerLys-----AlaThrIleGly 62
DB 333 TTCTCGGAGCGCTGTGTGAGTGTCTACACATGATGAACGAGAGATCCAGCGGAGGAC 392
QY 63 TyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleValValSer 82
DB 393 ATCGACCTGCCCTCACCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452
QY 83 ThrAspArgGlyLeuGlyGlyLeuAsnValLysLysValGlyMetIleValValSer 102
DB 453 GCGGAGCGCGCGCTCTCGGAGGCTTCAACACACGCTGCTCAAGAGCGGAAACCGCG 512
QY 103 MetLysGluThrLysGluLysAspValSerValGlnLeuSerLeuIleGlySerLysSer 122
DB 513 ATCGATGAGCTCAAGCAGCTGGCTCCAGTACACCTGCTCAGCGTGGGAGGAGGCGG 572
QY 123 IleAsnPheGlnSerLeuGlyLysLysLysLysLysLysLysLysLysLysLys 142
DB 573 AACGCTACTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 632
QY 143 ThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyrLysLys 162
DB 633 GTGCCACCGTCAAGGACTCGCAGGCGCATCTGGACCTGCTACTCTACTCTCTCTCTCT 692
QY 163 GlyGluValAspValTyrLeuValTyrAsnLysPheIleAsnThrMetSerGlnLys 182
DB 693 GAGGAGTGGCAAGGTGGAGTGTCTCTACTCAAGTGTGTGTGTGTGTGTGTGTGTGTGT 752
QY 183 ProValLeuGluLysLeuIleProLeuPro----- 192
DB 753 CCCATCATCAGAGCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 812
QY 193 -----GluLeuAspAsnAspGluLeu----- 199
DB 813 GTCTGCTGAGCGCCACCGAGGAGCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 872
QY 200 ---GlyGluArgLysGlnVal-----TyrAspTyrIle-----Tyr 210
DB 873 ACCGTGAGCGCGAGAGGTGAAGATCGAGACCGACCGCTCTCTCTCTCTCTCTCTCTCT 932
QY 211 GluProAspAlaLysValLeuLeuAspAsnLeuValArgTyrLeuGluSerGlnVal 230
DB 933 GAGCAGACCCCGTGCAGATCTCTGGACGGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 992

```

```

QY 231 TyrGlnAlaAlaValGluAsnLeuLaserGluGlnAlaAlaArgMetValAlaMetLys 250
DB 993 CTCCTGTCCTCAGGAGTGTCTGCCAGCGAGTCTGCCGCGCGATCAGCGCCATGAGC 1052
QY 251 AlaAlaThrAspAsnAlaGlyAsnLeuIleAsnGluLeuGlnLeuValTyrAsnLysAla 270
DB 1053 AGCGCCACCGACCAACGCGATCGAGTCTCGCAGAACCTCTCCATCGCGCTACACCGCAG 1112
QY 271 ArgGlnAlaSerIleThrAsnGluLeuAsnGluLeuValAlaAlaAlaIle 289
DB 1113 CGCCAGGCCAAGATCACCAGGAGATCTCGAGATCTCGCGGTGCGCGCGCTC 1169

CNSOA3T2 1313 bp mRNA linear HTC 06-FEB-2004
Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSLTPGH842B07 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).
BX828395
VERSION BX828395.1 GI:42461763
KEYWORDS HTC; GSLT cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (Bases 1 to 1313)
AUTHORS Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C.,
Menard,M., Cruaud,C., Quetier,F., Scarpetti,C., Schachter,V.,
Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
TITLE Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation
JOURNAL Unpublished
REFERENCE 2 (Bases 1 to 1313)
AUTHORS Genoscope.
JOURNAL Direct Submission
TITLE Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli
V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
URGV INRA, Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences). 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EP/Full
length
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES
source
Location/Qualifiers
1..1313
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
/db_xref="taxon:3702"
/clone="GSLTPGH842B07"
/tissue_type="Hormone Treated Callus"
/plasmid="pCMVSPORT6"
complement(1..1313)
/gene="At4g04640"

gene
ORIGIN
Alignment Scores:
Pred. No.: 3.47e-42 Length: 1313
Score: 440.00 Matches: 113
Percent Similarity: 53.58% Conservative: 59
Best Local Similarity: 35.20% Mismatches: 115
Query Match: 30.79% Indels: 34

```

DB: 3 Gaps: 5

US-09-545-199F-4 (1-289) x CNS0A3T2 (1-1313)

Qy 2 AlaGlyAlaLysGluLeuThrValSerValLysSerThrGlnLysIleThr 21  
 Db 128 GCCTCTCTCGAGCTCAGACCGTATCGATTGAGTCAAGGACACTCAAGATCACC 187

Qy 22 LysAlaMetGluMetValAlaLysMetArgLysThrGlnGluArgMetSer 41  
 Db 188 GAAGCTATGAAGCTGTCTGCTCAGCTAAGTCAAGGAGCTCAGAGCTGTGTCAAT 247

Qy 42 SerArgProTyrSerGluThrIleArgAsnValLysSerHisValSerLys 59  
 Db 248 GCACGACCATCTCAGAACCTTGTGAAGTCTTTACAACTCAACGACGACGTTCAA 307

Qy 60 ThrIleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIle 79  
 Db 308 ACCGATGATGTCGATGTTCTTAAACCAAGTCAGACCGTTAAGAAGTAGCTCTCGTT 367

Qy 80 ValValSerThrAspArgGlyLeuGlySerGlyLeuAsnValLysThrVal 99  
 Db 368 GTCGTCACCGGTATCGGTGATTAAGTGTGATTCACAACTTCATCAAGAAAGCA 427

Qy 100 LeuAsnGluMetLysGluThrLysGluLysAspValSerValGlnLeuSerLeuIleGly 119  
 Db 428 CAGGCAAGCAATCAAGAGCTTAAGGCTAGGCTTGAATACAGTCAATAGCGTGGC 487

Qy 120 SerLysSerIleAsnPhePheGlnSerLeuGlyLysLysLeuThrGlnAspSerGly 139  
 Db 488 AAGAAGGAATTTCTTCTCCGTCGCGCTATCATCCCGTCACAAATACCTAGAA 547

Qy 140 IleGlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspIle 159  
 Db 548 CGCGGAATTTACTACGGCTAAGAGAGCTCAAGCTGTGGCTGATGATGCTCTCTCTG 607

Qy 160 TyrLysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMet 179  
 Db 608 TTATAGTGAAGATCGAAAGTCGAGCTCTGTACACAACTTTGTATCTTGGTC 667

Qy 180 SerGlnLysProValLeuGluLysLeuIleProLeuPro 192  
 Db 668 AATCAGAACCCGTGATCTACAGCTACTGCTCTTATCACCTAAAGGAGATCTGTGAC 727

Qy 193 199  
 Db 728 ATTAATGGAACCTGTGTGGATGCTCGGAAGATGAATTTTCAGTTAAACGACAAAGAA 787

Qy 200 GlyGlu 209  
 Db 788 GGGAAATTCACAGTTGAAGACAGACTTTTAGGACACCAACAGCTGATTTCTCGCCGATC 847

Qy 210 227  
 Db 848 TTGCAATTCGAGCAAGACCTGTTCAGATCTTGTGCTTTGTGCTCTGTATCTTAAC 907

Qy 228 SerGlnValTyrGlnAlaValGluAsnLeuLysSerGluGlnAlaLysMetVal 247  
 Db 908 AGTCAGATCTTAGGCGATACAGAGTCAATGTGCTAGTGTGAGCTTGCAGTAGAATGAT 967

Qy 248 AlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIleAsnGluLeuGlnLeuValTyr 267  
 Db 968 GCAATGAGTAGTGTCTCGGATATGATCGATCTCAAGAAATCGCTTTCGATGGTGTAT 1027

Qy 268 AsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsnGluIleValAlaGlyAlaAla 287  
 Db 1028 AATGAAAGCGTCAAGCTAAGATTACTGGAGAGATTTCTTGAGATTTGCTCGAGCTAAT 1087

Qy 288 Ala 288  
 Db 1088 GCA 1090

RESULT 6

BH387684 807 bp DNA linear GSS 11-DEC-2001  
 AG-ND-167A14.TR ND-TAM Anopheles gambiae genomic clone  
 ACCESSION BH387684  
 VERSION BH387684.1 GI:17333825  
 KEYWORDS GSS.  
 SOURCE Anopheles gambiae (African malaria mosquito)  
 ORGANISM Anopheles gambiae  
 Eukaryota; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles 1 (bases 1 to 807)  
 REFERENCE Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J. and Collins,F.H.  
 TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae  
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)  
 MEDLINE 22542063  
 PUBMED 12655398  
 COMMENT Other GSSs: AG-ND-167A14.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org  
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
 Seq primer: M13 Rev  
 Class: BAC ends.  
 FEATURES  
 Location/Qualifiers  
 1..807  
 /organism="Anopheles gambiae"  
 /mol\_type="genomic DNA"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-167A14"  
 /clone\_lib="ND-TAM"  
 /note="Vector: pSCBAC1; Site 1: HindIII"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,41e-41 Length: 807  
 Score: 432.50 Matches: 102  
 Percent Similarity: 62.13% Conservative: 67  
 Best Local Similarity: 37.50% Mismatches: 95  
 Query Match: 30.27% Indels: 8  
 DB: 4  
 Gaps: 4

US-09-545-199F-4 (1-289) x BH387684 (1-807)

Qy 19 LysAlaThrLysAlaMetGluMetValAlaLysSerLysMetArgLysThrGlnGluArg 38  
 Db 803 CAATACACAGTCTATGAAATGTTTCGCGAGCGAACCTTAAGAAAGACACAGATGCT 744  
 Qy 39 MetSerSerSerArgProTyrSerGluThrIleArgAsnValLysSerHisValSerLys 58  
 Db 743 ATTGTGATGCTTCGCCCTTACTCCGAAAACTTCAGGAATAATAGAAAAACGTTAGT-- 686  
 Qy 58 sAlaThrIleGlyTyrLysHisProPheLeuValAsp-----ArgGluValLysLys 75  
 Db 685 -TCTACTAGATCAGGAGACGCTGCTGTGTGTTGATCGACACACAGGAGGTTAAAG 627





This clone is from an *A. gambiae* BAC library (ND-TAM) provided by P. H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center, University, College Station, Texas 77843-3123, USA using a HindIII partial digest.

Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
SOURCE

```

location/Qualifiers
1..796
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7155"
/clone="AG-ND-142J4"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; site 1: HindIII"

```

## ORIGIN

Alignment Scores:		
Pred. No.:	5.25e-38	Length:
Score:	403.00	Matches:
Percent Similarity:	60.00%	Conservative:
Best Local Similarity:	3.55%	Mismatches:
Query Match:	28.2%	Indels:
DE:	8	Gaps:
		956
		796

US-09-545-199F-4 (1-289) X BH393596 (1-796)

Oy      20 IleThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMet 39  
      :: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db      796 TTACCAGTGTATGAATAATGGTTCCGGCAGCGAARACTAAAGAARGACACAGATGCTATT 737

QY 40 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerIys--- 58

Db  
736 GTGATGCTTCGCCCTTACTCGAAAACTTCAGGAAATAATAGAAACGTTAGTTCTACA 677

**OY**      ---AlaThrIleGlyTyrLysHisProPheLeuValAspArgGluVal 73  
             |||   :|||  
**Ddb**      676 CTAGATCAGGAACCGCTTCGGTT---TATGCACAACCA---ACAGAAGTT 632

74<sup>1</sup> LysLysValGlyMetIleValValSerThrAspArgGlyIleuCysGlyGlyIleuGlnVal 83  
-----AGAGAGGTT 633

74 MySvsvagsglyMeullelvalwalserinnrAspArgglyLeuCysGlyGlyLeuAanVal 93  
::: ::  
631 AAAAGAACTCTGTTCATTGCAGTTACTTCCACACAGAGTCCTTGCAAGAGCTTTCAACTCT 572

D8 Db

94 AsnLeuPheLysThrValLeuAsnGluMetLysGluTrpLysGluLysAspValSerVal 113

Db  
571 TCTGTGATATCAAGAA-GTTAATGCTCAGTATCAG--CAAAATTCAGCTTTTGAGGTT 516

DY  
114 GlnLeuSerLeuIleGlySerLysserIleAsnPhePheClnSerLeuGlyLeLysile 133

: : : : :  
: : : : :  
: : : : :

b  
515 GAAGTGTGACTATTGGTAAAAAGCATTGATGCTTTCAGAGCT---TCTAAGAAATC 459

[illegible]

b  
458 TATGATAACCACACGGATCTTTACGATCTTTTAACTTTGAAGCGTAGCATGTAGCT 399

[illegible]

174	LYSPHEILLEASNTHRME	SERGLN	LYSPROVAL	LEUGLU	LYSLEU	ILEPRO	LEUPROGLU	193
339	CAAGGCAATCATGACAG	ATTTCCG	CACGGGAGT	TTTTTGT	GATGAAG	ATATACCT	TGCTGTACAAC	339

b  
338 AAATTCCTTAATCGCGCAACGCAGGAAGTACACAGAAAAGCTTCTTCCTATTACAATG 279

194 LeuAspAsnAspGluLeuGlyGluArgLysGlnValTyrAspTyrIleTyrGluProAsp 213

b  
278 CCTGAAAGAAAGAGCTGAGAGAGTGCATCGAAACCGATTATATCTTCGAACCAAC 219

Y 214 AlaLysValLeuLeuAspAsnLeuLeuValArgTyrLeuGluSerGlnValTyrGlnAla 233

Db	218	CGTACCGAAATTTCTGGAAACCTTAAATCTCTCAAGACTCAGGTTTATAAGCG	159
Qy	234	AlaValGluAsnLeuAlaSerGluGlnAlaAlaArgMetValAlaMetLysAlaAlaThr	253
Db	158	ATCCTTGATTCAGTAGCTTCTGAGCATGGTGCAAGGATGACTGCAATGCACAAAGCAACA	99
Qy	254	AspAsnAlaGlyAsnLeuIleAsnGluLeuGlnLeuValTyraAsnLysAlaArgGlnIle	273
Db	98	GATAATGCACAGGCATTCGAAGAATGACCTGGTTATCTTCTACAAACAAGCAGCTCAGGCT	39
Qy	274	SerIleThrAsnGluLeuAsnGluIleVal	283
Db	38	GCAATTACCAACGAGACTCTTGAAATTGATC	9
RESULT	9		
LOCUS	BZ569821/c		
DEFINITION	msH2_1034.y2 msh Pseudomonas aeruginosa genomic clone msh2_1034,	1474 bp	DNA linear
ACCESSION	BZ569821		GSS 17-DEC-2002
VERSION	BZ569821.1		
KEYWORDS	GI:27204882		
	GSS.		

ORGANISM	Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.	
REFERENCE	1 (bases 1 to 1474)

AUTHORS	TITLE
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burris, J.L., Kaul, R. and Olsen, M.V.	Whole-Genome-Sequence variation among multiple isolates of <i>Pseudomonas aeruginosa</i> lineage

JOURNAL  
COMMENT  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond

CONTACT: CHRIS K. RAYMOND  
Genome Center  
University of Washington

Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2062216954

**Fax: 2066857244**  
**Email: [craymond@.washington.edu](mailto:craymond@.washington.edu)**  
**Class: shotman**

```

FEATURES
source
Location/Qualifiers
1. .1474
Class: shotgun.

```

```

source
1. .1474
/organism="Pseudomonas aeruginosa"
/mol type="genomic DNA"

```

```

/mori_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"

```

```

/clone="msh2_1034"
/clone_lib="msh"

```

```
ORIGIN
/note="Environmental isolate. Whole genomic shotgun
library."
```

Alignment Scores:

Pred. No.:	5.94e-35	Length:	1474
Score:	381.00	Matches:	72

Percent Similarity:	84.40%
Best Local Similarity:	66.06%
Query Match:	32.50%
Mismatches:	17
Conservative:	20
Mismatches:	72

```

Query Match: 26.66%
DB: 8
Indels: 0
Gaps: 0

```

OS-09-545-199F-4 (1-289) x BZ569821 (1-1474)

1 MetAlaGlyAlaLysGluLeuArgThrLysIleAlaSerValLysSerThrGlnLysIle 2

432 ATGGCAGGCGCAAAAGAGATTTCGAGCAAGATTTCGAGCATCAAAAGCAGCAAAAGATC 3

**a**

ThrLysAlaMetGluMetValAlaIleSerLysMetArgLysThrGlnGluArgMetSer  
21 372 ACCAATGCCATGGCAAAAAGTCCCGCTGTCCACATGTCGAAAGCAGGCAGG

3/2 ACCAAATGCCATGGAAAAAGTGGCGGTGAGCAAGATCGCGAGGCCACAAATGCCATGGCG 41 SerSerArgProTyrSerGluThiIleArgAsnValIleSerHisValSerIleValAlaIle

[illegible]

```

Db 312 GCCGCGCTCCCTACCGGAGCGTATTTCGCCAGGTGATCGGCATCTGCGCCAAOCGCAAC 253
Qy 61 lIedGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Db 252 CCGGAATACCGCTACCCCGTTTCATGCTCGAGCGTGAAGTCAAGCGCTGCCTACATCGTG 193
Qy 81 ValSerThrAspArgGlyLeuCyGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db 192 GTGAGCAGTACCGCTGCTGCTGCGCGGCTGCTGACATCAACCTGTTCAAGTCCCTCGTC 133
Qy 101 AsnGluMetLysGluTrpLysGluLys 109
Db 132 AAGGACATGAGCGCTACCGCGCA 106

RESULT 10
LOCUS AQ936809/c
DEFINITION NB1-026R Human NotI clones/Homo sapiens genomic, genomic survey
sequence.
ACCESSION AQ936809
VERSION AQ936809.1 GI:7213187
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 448)
AUTHORS Zabarovsky,E.R., Gizatullin,R., Podowski,R.M., Zabarovsky,V.V.,
Xie,L., Muravenko,O.V., Kozirev,S., Petrenko,L., Skobeleva,N.,
Li,J., Protodopov,A., Kashuba,V., Ernberg,I., Winberg,G. and
Wahlstedt,C.
TITLE NotI clones in the analysis of the human genome
JOURNAL Nucleic Acids Res. 28 (7), 1635-1639 (2000)
MEDLINE 20175728
PUBMED 10710430
COMMENT Contact: Podowski RM
Center for Genomics Research
Karolinska Institute
17177 Stockholm, Sweden
Tel: +46-8-728-6372
Fax: +46-8-337583
Email: Raf.Podowski@cgr.ki.se
Class: NotI site.
FEATURES
    source
    1..448
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /clone_lib="Human NotI clones"
ORIGIN
Alignment Scores:
Pred. NO.: 1.64e-35 Length: 448
Score: 379.50 Matches: 85
Percent Similarity: 70.27% Conservative: 19
Best Local Similarity: 57.43% Mismatches: 37
Query Match: 26.56% Indels: 7
DB: 8 Gaps: 2

US-09-545-199f-4 (1-289) x AQ936809 (1-448)
Qy 144 ProSerValGluInLeuIleGlySerValAsnSerMetIleAspAla-TyrLysLysG 163
Db 433 CCCACATCGAGCGCCGTGATCGGCCCGCTCAAGTGTCTGTCGAGCCNCTACGCCGAGG 374
Qy 163 yGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSerGlnLysPr 183
Db 373 CAGGCTGTGCGCGTCTACTCTGCTACCCAGTTCATCACAGATGACGATGACGAGGT 314
Qy 183 oValLeuGluLys-LeuIleProGluProGluLeuAspAsnAspGluLeuGlyGluArg 203
Db 313 GCAGGTGCACCGCTCTGCTGCGCTCAAGGCC-----AGCGAGTGGCGAGGAA 263

```

```

Qy 203 ySgInVal-----TTPAspTyrLleTyrGluProAspAlaLysValLeuLeuAspAsnL 221
Db 262 AGCGCGCCACTCGTGGGATACATCTACGAGCCGATGCCNCGACNGTCATCATCGAAC 203
Qy 221 euleuValArgTyrLeuGluSerGlnValTyrGlnAlaValAlaValGluAsnLeuLeuLeu 241
Db 202 TGTGTGTGGCTACTCGGAAGCGTGTCTTCCAGCGCGTCCGAGACATGCGCGAGCG 143
Qy 241 luGlnAlaAlaArgMetValAlaValMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLeu 261
Db 142 AGCASTCCGCGCGCATGTGGCCATGAAGGTGGACCGNCAACGCGCGCGCATGTCG 83
Qy 261 snGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsnG 281
Db 82 GCGAGCTGAAGCTGTCTACANCAAGACCCGCCAGCGCGCATCACCAAGAGTTGTCCG 23
Qy 281 luileValAlaGluValAla 287
Db 22 AGATGCTCAGCGGCGCGGCC 3

```

```

RESULT 11
LOCUS BH378165/c
DEFINITION

```

```

AG-ND-138P6.TF ND-TAM Anopheles gambiae genomic clone AG-ND-138P6,
genomic survey, sequence.
ACCESSION BH378165
VERSION BH378165.1 GI:17324307
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae

```

```

REFERENCE 1 (bases 1 to 666)
AUTHORS Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
Gardner,M.J. and Collins,F.H.
TITLE Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)
MEDLINE 12655398
PUBMED
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org

```

```

This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
FEATURES
    source
    1..666
    /organism="Anopheles gambiae"
    /mol_type="genomic DNA"
    /strain="PEST"
    /db_xref="taxon:7165"
    /clone="AG-ND-138P6"
    /clone_lib="ND-TAM"
    /note="Vector: pECBAC1; Site_1: HindIII"
ORIGIN
Alignment Scores:

```

```

Pred. No.: 5.48e-32 Length: 666
Score: 352.50 Matches: 80
Percent Similarity: 61.19% Conservative: 54
Best Local Similarity: 36.53% Mismatches: 82
Query Match: 24.67% Indels: 3
DB: 8 Gaps: 2

US-09-545-199F-4 (1-289) x BH378165 (1-666)

QY 70 AspArgGluValLysValGlyMetIleValSerThrAspArgGlyLeuCysGly 89
Db 648 GAGAGTAGGTAAAGAAATCCTGTCATTCGAGTCATTCACAGAGAGTCTTCAGGA 589
QY 90 GlyLeuAsnValLeuLeuPheLysThrValLeuAsnGluMetLysGluTrpLysGluLys 109
Db 588 GCTTTCAACTCTCTGTATCAAGAAGTTAATGCTCAGTATCAGCAA-----AATCA 535
QY 110 AspValSerValGluLeuSerLeuLeuGlySerLysSerIleAsnPhePheGlnSerLeu 129
Db 534 GCTTTGAGGTTGAGGTTGAGTATTTGTAACACAGCAATTTGATGCTTCAGAGCT--- 478
QY 130 GlyLeuLysLeuThrGlnAspSerGlyIleGlyAspThrProSerValGluGlnLeu 149
Db 477 TCTAAGACATCTATGATAACACAGCATCTTACGATCTTTAACTTTGAAGCGTA 418
QY 150 IleGlySerValAsnSerMetIleAspAlaTyrLysLysGlyGluValAspValTyr 169
Db 417 GCACATGTAGTCAGGATATCATGAGAGATTTCCGTCAGGGAGTTTTTGTGATGAAGTATAC 358
QY 170 LeuValTyrAsnLysPheIleAsnThrMetSerGlnLysProValLeuGluLysLeuIle 189
Db 357 CTGGTGTACAAACAATTCCTTAATCGGCAACGACGAGGAGTACACAGAAAACCTTCTT 298
QY 190 ProLeuProGluLeuAspAsnAspGluLeuGlyGluArgLysGlnValTrpAspTyrIle 209
Db 297 CTTATTCAATCGCTGAAAGAAAGAGCTGAGAGAGTGCATCGAACCGATTATATC 238
QY 210 TyrGluProAspAlaLysValLeuLeuAspAsnLeuLeuValArgTyrLeuGluSerGln 229
Db 237 TTGCAACCAAAACCGTACCGAAATCTGGAAACCTTAATCTCTAATCTATCAAGCTCAG 178
QY 230 ValTyrGlnAlaValGluAsnLeuLysGluGlnAlaAlaArgMetValAlaMet 249
Db 177 GTTTATAGCGATCTCTGATTTCAGTATCTCTGAGAAAAGCAAGGATGCTGCAATG 118
QY 250 LysAlaAlaThrAspAsnAlaGlyAsnLeuIleAsnGluLeuGlnLeuValTyrAsnLys 269
Db 117 CACAAGCAACAGATAATGCACAGCATTTGAAGATGACCTGGTATCTTCTTCAACAAA 58
QY 270 AlaArgGlnAlaSerIleThrAsnGluLeuAsnGluIleValAlaGlyAlaAlaAla 288
Db 57 GCACGTGAGGCTGCAATTACCAACAGATCTTGAAATTTGATCCGAGCAGAGACT 1

RESULT 12
LOCUS CNS0A9BS 1205 bp mRNA linear HTC 06-FEB-2004
DEFINITION Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSUTLS16ZE03 of Adult vegetative tissue of strain col-0 of
Arabidopsis thaliana (thale cress).
BX819675
GI:42468345
VERSION HTC; GSUT cDNA.
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1205)
Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C.,
Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome

```

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Annotation  
Unpublished  
2 (bases 1 to 1205)  
Genoscope.  
Direct Submission  
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT

The sequences are based on single pass reads.  
Life Technologies (a division of Invitrogen) members carried out  
full-length libraries construction : Temple G.  
Genoscope members carried out sequencing and annotation : Castelli  
V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,  
Schachter V., Weissenbach J., Salanoubat M.  
URGV INRA : Clepet C., Caboche M.  
Annotation is based on the June 2003 version of the Arabidopsis  
genome released by MIPS (Munich Information center for Protein  
Sequences). 5 prime and 3 prime are assembled with Phrap.  
http://www.genoscope.cns.fr/externe/sequences/Banque\_Projet\_EF/Full  
length  
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.

FEATURES

source

1..1205  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/strain="Col-0"  
/db\_xref="taxon:3702"  
/clone="GSUTLS16ZE03"  
/tissue\_type="Adult vegetative tissue"  
/plasmid="pCMVSPORT 6"  
complement(1..1205)  
/gene="At2g33040"

gene

ORIGIN

Alignment Scores:

Pred. No.: 2.24e-31 Length: 1205  
Score: 350.50 Matches: 106  
Percent Similarity: 49.84% Conservative: 50  
Best Local Similarity: 33.87% Mismatches: 88  
Query Match: 24.53% Indels: 69  
DB: 3 Gaps: 12

US-09-545-199F-4 (1-289) x CNS0A9BS (1-1205)

```

QY 7 IleArgThrLysIleAlaSerValLysSerThrGlnLysIleThrLysAlaMetGluMet 26
Db 245 GTGGGTAAACCGCATGAAGAGTGTTAAGACATCAAAAGATCAAAAGGCAATGAAGATG 304
QY 27 ValAlaAlaSerLysMetArgLysThrGlnGluArgMetSerSerArgProTyrSer 46
Db 305 GTTCTGCTTCCAGCTTAGAGCAGTTCAGAGCGGCGAGCTGAGAACTCCCGT----- 355
QY 47 GluThrIleArgAsnValIleSerHisValSerLysAlaThrIleGlyTyrLysHisPro 66
Db 356 -----LeuValAspArg-----GluValLysValGlyMetIleVal 370
QY 67 Phe-----LeuValAspArg-----GluValLysValGlyMetIleVal 80
Db 371 TTACTGCATCTTAGGAGATATCCAGCATTTAGTAAAGAGAGTGTGTGTGTCCT 430
QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThr----- 98
Db 431 CTCTCTTGTGACAAGGGTCTCTGTGTGTGAATCACTCCACTGTCTGTTAAAGTAGAGCAGG 490
QY 99 ValLeuAsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIle 118
Db 491 GCTCTGTACAAATTTGAATCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544
QY 119 GlySerLysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSer 138
Db 545 GGAGAGAAAGCA-----AAGGCTATAATGTTTCGTGACTCA 580
QY 139 GlyIleGlyAspThrProSerValGluGlnLeu-----IleGlySer 152

```

```

Db 581 AAGAACGACATGTGCTCTCTGTAACAGAGCTGAATAAGAACCCACTCAATATGCTCAG 640
Qy 153 ValAsnSerMetIle---AspAlaTyrLysGlyGluValAspValValTyrLeuVal 171
Db 641 GTGTCACTTCTAGCTGATGACATCTCGAAGACGTTGAATTTGATGCTTTGCGCATTTGC 700
Qy 172 TyAsnLysPhe-----IleAsnThrMetSerGlnLys 182
Db 701 TACAACAAGTTCATTTCAGTTGCGCATTTCTGCCAACTGTGTCCACTGTTTGTCACT 760
Qy 183 ProValLeuGluLysLeuLe---ProLeuProGluLeuAspAsnAspGlu 198
Db 761 GAGATTATTGAGAGAGAGCTCGAATTTGGAGAAACTTGGCGAGCTTGACTCATATCAG 820
Qy 199 Leu-----GlyGluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysVal 216
Db 821 ATTGAAGTGGGGGAACAAG-----GGAGAA 847
Qy 217 LeuLeuAspAsnLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGlu 236
Db 848 ATATTGCAAGACTTGGCGAGTCCAAATCTCTGTGTGATGTTTCAATGCGGTTCTGGAG 907
Qy 237 AsnLeuAlaSerGluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAla 256
Db 908 AATGCATGTAGTGAGATGGAGCAAGATGTCGCCATGGACAGCTCAACGAAACGCA 967
Qy 257 GlyAsnLeuLeuAsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThr 276
Db 968 GGAGAAATGCTCGACCGCTCACCTCATACACAGAGACTCGTCAAGCTTCTATTACA 1027
Qy 277 AsnGluLeuAsnGluLeuValAlaGlyAlaAlaLaille 289
Db 1028 ACAGAGCTTATTGAGATTCTCTGGAGCTTCTCTCTT 1066

RESULT 13
LOCUS BH385025/c
DEFINITION AG-ND-167A6/TR ND-TAM Anopheles gambiae genomic clone AG-ND-167A6,
ACCESSION BH385025
VERSION BH385025.1 GI:17331167
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 667)
Hong, Y. S., Hogan, J. R., Wang, X., Sarkar, A., Sim, C., Loftus, B. J.,
Rep, C., Huff, E. R., Carlisle, J. L., Black, K., Zhang, H. B.,
Gardner, M. J., and Collins, F. H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
12655398
Other_GSSs: AG-ND-167A6.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center

```

University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
Location/Qualifiers  
1..667  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strains="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-167A6"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.27e-31 Length: 667  
Score: 349.50 Matches: 85  
Percent Similarity: 60.71% Conservative: 51  
Best Local Similarity: 37.95% Mismatches: 81  
Query Match: 24.46% Indels: 8  
DB: 3 Gaps: 3

US-09-545-199F-4 (1-289) x BH385025 (1-667)

Qy 63 TyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleValValSer 82  
Db 651 TATCCACCAACA-----AGAGAGGTAAGAAATCTGTTCATTGCAGTTACT 604  
Qy 83 ThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeuAsnGlu 102  
Db 603 TCCAAACAGAGGCTTCGAGGAGCTTCAACTCTCTGTTTATCAAGAGCTTAAATGCTCAG 544  
Qy 103 MetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySerLysSer 122  
Db 543 TATCAGCAA-----AATTCAGCTTTTGGGTGAAGTTGACTATTGGTAAAAAGCA 490  
Qy 123 IleAsnPhePheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIleGlyAsp 142  
Db 489 TTTGATGCTTTTCAGAGCT--TCTAAGAAATCTATGATAACACACAGGAGCTTTTACGAT 433  
Qy 143 ThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyrLysLys 162  
Db 432 TCTTTAACTTTTGAAGCGGTAGACATGATGCTGAGGATATCATGAGAGATTTCGTCAG 373  
Qy 163 GlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSerGlnLys 182  
Db 372 GGAGTTTTTGTGATGAAGTATACCTGGTGTACACAAATTCCTTAAATCGGCAACGAGGAA 313  
Qy 183 ProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGlyLysArg 202  
Db 312 GTACACAGAAAGAGCTTCTTCTTATTACAAATGCTGAAAGAAAGAAAGCTGAGAGAGT 253  
Qy 203 LysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsnLeuLeu 222  
Db 252 GCAATCGAAACCGATTATATCTTCGAAACCAACCGTCCCGAAATTCGGAACCTTAATT 193  
Qy 223 ValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuLaserGluGln 242  
Db 192 CTTAAATCTATCAGAGACTCAGGTTTATAAAGCGATCCTTGATTTCAGTAGCTTCTGAGCAT 133  
Qy 243 AlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIleAsnGlu 262  
Db 132 GGTGCAAGGATGACTGCAATGCACAAAGCAACAGATAATGCACAGGATTGGAAGATGAC 73  
Qy 263 LeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGlnLeuAsnGluIle 282  
Db 72 CTGGTTATCTTCTCAACAAAGCAGCTCAGGTCGATTCATTCACCAACGAGATCTT-GAAATT 14  
Qy 283 ValAlaGlyAla 286  
Db 13 GTATCCGGAGCA 2

Qy	30	SerLysMetArgLysThrGlnGluArgMetSerSerSerArgProTyrSerGluThrIle	49
Db	63	TCCAAGCTTAGACAGCTTCAAGCCGAGCTGAGAATCCCGT	104
Qy	50	ArgAsnValIleSerHisValSerLysAlaThrIleGlyTyrLysHisProPhe	67
Db	105	-----GGACTTTGGCAGCCATTTACTGCA	128
Qy	68	---LeuValAspArg	---
Db	129	CUTCTAGGAGATAATCCACAGATTGATGTAAGACAGAGTGCTGTCTCTCTCT	188
Qy	84	AspArgGlyLeuCyssGlyLeuAsnValAsnLeuPheLysThr	101
Db	189	GACAGGGTCTCTGTGGTGAATCAACTCCACTGTCGTTAAAGTGAGCAGGCTCTGTAC	248
Qy	102	GluMetLysGluTrpLysGlyLysAspValSerValGlnLeuSerLeuIleGlySerLys	121
Db	249	AAATTGAATGCTGGTCTGAAAAGGAA-----GTTCAAGTTGTTATTGTGGAGAGAAA	302
Qy	122	SerIleAsnPheGlnSerLeuGlyLysIleLeuThrGlnAspSerGlyIleGly	141
Db	303	GCA-----AAGGCTAATATGTTTCGTGACTCAAAGAACAC	338
Qy	142	AspThrProSerValGluGlnLeu-----IleGlySerValAsnSer	155
Db	339	ATTGTCTCTCTGTAACACAGACTGAATAAGAACCCACTCAATTATGTCAGGTGTCAATT	398
Qy	156	MetIle---AspAlaTyrLysGlyGluValAspValValTyrLeuValTyrAsnLys	174
Db	359	CTAGCTGATGATCCTCTGAAGAACGTTGCAATTTGATGCTTTGGCATTGTCTACACAAG	458
Qy	175	Phe-----IleAsnThrMetSerGlnLysProValLeu	185
Db	459	TTCATTTCAGTTGCGCATTTCTGCCAACGTGTCCACTGTTTGTCCACTGAGATTATT	518
Qy	186	GluLysLeuIle-----ProLeuProGluLeuAspAsnAspGluLeu	199
Db	519	GAGAAGAGCTCTGAAATGGAGGAAACTTGGCAGAGCTTGACTCATATGAGATTGAAGT	578
Qy	200	GlyGluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAsp	219
Db	579	GGGGAAACAAG-----GGAGAAATATTGCG	605
Qy	220	AsnLeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAla	239
Db	606	AATCTGCCAGTCCAAATCTCTGTGTGATGTTCAATGCCGTTCTCGAGAATGCATCT	665
Qy	240	SerGluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeu	259
Db	666	AGTGAGATGGAGCAAGAATGCTGCCATGGACAGCTCAAGCAGAAACCCAGAGAAATG	725
Qy	260	IleAsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeu	279
Db	726	CTCGACCGTCTCACCTCACATACACAGAGACTCGTCAAGCTTCTATTACACAGAGCTT	785
Qy	280	AsnGluIleValAlaGlyAlaAlaIle	289
Db	786	ATTGAGATTATCTCTGGAGCTTCTCTCT	815
RESULT	15		
LOCUS	AY108441		
DEFINITION	Zea mays PC0083558 mRNA sequence.	1:14 bp	linear
ACCESSION	AY108441		
VERSION	AY108441.1	GI:21211519	
KEYWORDS	HTC.		
SOURCE	Zea mays		
ORGANISM	Zea mays		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 1514)		

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: November 9, 2004, 07:55:18 ; Search time 639 Seconds  
(without alignments)  
2435.070 Million cell updates/sec

Title: US-09-545-199F-4  
Perfect score: 1429  
Sequence: 1 MAGAKEIRTKIASVKSTQKI.....ARQASITNELNEIVAGAAAI 289

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 3611042 segs, 2692057975 residues  
Total number of hits satisfying chosen parameters: 7222084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xl  
-O=/cgn2\_1/USPTO\_pool\_p/US09545199/runat\_08112004\_112527\_21010/app\_query.fasta\_1.455  
-DB=Published Applications NA -QMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl6sum62  
-TRANS=human40.cdi -L-ST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09545199 @CNG 1.1 912 @runat 08112004\_112527\_21010  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1429	100.0	1972	11	US-09-809-665A-3
C 2	1173	82.1	1830121	14	US-10-329-960-1
C 3	1173	82.1	1830121	16	US-10-329-960-1
C 4	1173	82.1	1830121	18	US-10-158-865-1
5	1132.5	79.3	867	11	US-09-809-665A-132
6	1077.5	75.4	866	11	US-09-809-665A-166
7	1007	70.5	861	15	US-10-369-493-44931
8	1002	70.1	861	15	US-10-369-493-44931
9	991	69.3	864	9	US-09-741-669-144
10	991	69.3	864	15	US-10-369-493-47300
11	904.5	63.3	858	15	US-10-369-493-37300
12	891	62.0	867	15	US-10-369-493-32530
13	885.5	62.0	882	15	US-10-369-493-43286
14	885	61.9	94750	16	US-10-672-787-38
15	838	58.6	861	15	US-10-369-493-39109
16	838	58.6	861	15	US-10-369-493-39477
17	838	58.6	861	15	US-10-369-493-39477
18	812.5	56.9	876	15	US-10-369-493-39859
19	812.5	56.9	876	15	US-10-369-493-39859
20	809	56.6	861	15	US-10-369-493-30933
21	804	56.3	640681	9	US-10-369-493-33058
22	795	55.6	864	15	US-09-790-988-1
23	795	55.6	2731748	17	US-10-369-493-41252
24	787	55.1	861	15	US-10-297-465A-1
25	650.5	45.5	531	11	US-09-809-665A-97
26	585.5	41.0	843	15	US-10-369-493-32966
27	567.5	39.7	879	15	US-10-369-493-33589
28	536.5	37.5	858	15	US-10-369-493-32600
29	534.5	37.4	945	15	US-10-369-493-41141
30	530.5	37.1	864	15	US-10-369-493-43949
31	530.5	37.1	7750	9	US-10-369-493-47016
32	529	37.0	858	15	US-03-070-327A-216
33	524.5	36.7	4815	15	US-10-369-493-40239
34	523.5	36.6	870	15	US-10-310-630-1
35	518	35.7	851	9	US-09-974-300-100
36	509.5	35.7	945	15	US-10-369-493-43628
37	504.5	35.3	945	15	US-10-369-493-28299
38	504.5	35.3	2207	15	US-10-310-630-6
39	502	35.1	882	15	US-10-369-493-43236
40	501.5	35.1	888	17	US-10-474-776-529
41	501.5	35.1	10399	8	US-08-961-527-160
42	501.5	35.1	10399	16	US-10-158-844-160
43	489	34.2	684707	16	US-10-398-221-9
44	489	34.2	3011208	16	US-10-398-221-2058
C 45	486	34.0	6309	8	US-08-781-986A-480

ALIGNMENTS

RESULT 1  
US-09-809-665A-3  
; Sequence 3, Application US/09809665A  
; Publication NO. US20040110288A1  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David, et al.  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/00435  
; CURRENT APPLICATION NUMBER: US/09/809,665A  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 09/545,199  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1972  
; TYPE: DNA





```

/ LOCATION: (45593)..(45593)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (45732)..(45732)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (47036)..(47036)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (51334)..(51334)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (51602)..(51602)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (51786)..(51786)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (51805)..(51805)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (55369)..(55369)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (65309)..(65309)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (65313)..(65313)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (80024)..(80024)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (100091)..(100091)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (102696)..(102696)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (105121)..(105121)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (107248)..(107248)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (117136)..(117136)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (119750)..(119750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (119924)..(119924)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (120038)..(120038)

```

```

/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (122167)..(122167)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (122336)..(122336)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (131340)..(131340)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (131360)..(131360)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (139910)..(139910)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (140398)..(140398)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (142750)..(142750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152530)..(152530)

```

## Alignment Scores:

Pred. No.:	7,62e-108	Length:	1830121
Score:	1173.00	Matches:	221
Percent Similarity:	91.00%	Conservative:	42
Best Local Similarity:	76.47%	Mismatches:	26
Query Match:	82.09%	Indels:	0
DB:	14	Gaps:	0

US-09-545-199f-4 (1-289) x US-10-329-960-1 (1-1830121)

QY

1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerVallySerThrGlnLysIle 20

Db 501934 ATGGCAGGTGCAAAAGATATAAAACCAAAATGCCAGTGTACAAAGTACACAAAAATC 501875  
Qy 21 ThrlyAlaMetGluMetValAlaalaSerlyMetArglyThrGlnGluArgMetSer 40  
Db 501874 ACTAAGGCAATGGAATGCTGCAACCTCGAATATGCTAAAACGAGGATCGTATGGT 501815  
Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerlyAlaThr 60  
Db 501814 GCATCTCTCCGTATTCGAACTATCCGTAAGCTTATTAGTCATGTCTAAGCAAGT 501755  
Qy 61 lIleGlyTyrIleHisProPheLeuValAspArgGluVallyIleVallyMetIleVal 80  
Db 501754 ATCGTTTATAACATCGTTCTTCTTAGTTGACGGGAGAGTGNAGAAATCGGTATCTGGTT 501695  
Qy 81 ValSerThrAspArgGlyLeuCysGlyLeuAsnValAsnLeuPheIleThrValLeu 100  
Db 501694 ATTTCAACACATCGTGGATGTGTGGTGGTAAATGTTAATTTATTCACAAACACACTT 501635  
Qy 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerIleGlySer 120  
Db 501634 AACCAATAAAAAATGGAAAGAACAAATATTTCTACAGATTGGGCTTAATAGGTTCA 501575  
Qy 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140  
Db 501574 AAAGGATTAGTTTTTCGCTTCCTTGGATTATATCAAAAGTCAGCTTCTGGTTA 501515  
Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160  
Db 501514 GCGCATACCCGCTCTAGAACAAATTAATGTGTGGCAATACATATGTTGTGCTTAT 501455  
Qy 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
Db 501454 CGTAATGGTGAATGATGACGATTTATTTGTCATACATAAATTTGTTAATACGATCTG 501395  
Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
Db 501394 CAAAGACCTGTTGTACAACTATGCTCTTACAGAACTAAAGACGATCTTTAAAT 501335  
Qy 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
Db 501334 GAAAGACACACAGCTGGGATTTATCTTATGAGCAGACAAAGTACTATTAGATAGC 501275  
Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240  
Db 501274 CTTTATGTTCTGTTTATGATCTCCAAATTTATCAAGCGTTGTAGATATGTAGCTTCA 501215  
Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260  
Db 501214 GAAACAGCGCTCGAATGGTAGCAATGAAAGCAGCAGCGATATCGAGGTATTTAAT 501155  
Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
Db 501154 AATGATCTGCGTTGGTGTACACAAAGCTCGTCAAGCAAGTATCACAATGAATGAAT 501095  
Qy 281 GluIleValAlaGlyAlaAlaAlaIle 289  
Db 501094 GAAATCTAGCGGTGCGGCGGCAAT 501068  
RESULT 3  
US-10-329-670-1/c  
; Sequence 1, Application US/10329670  
; Publication No. US20040018503A1  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag  
; FILE OF INVENTION: Thereof, and Uses Thereof  
; FILE REFERENCE: PB186P1  
; CURRENT APPLICATION NUMBER: US/10/329,670  
; CURRENT FILING DATE: 2002-12-24  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (40810)..(40810)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (44416)..(44416)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (44905)..(44905)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (44975)..(44975)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (45593)..(45593)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (45732)..(45732)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (47036)..(47036)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (51334)..(51334)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (51602)..(51602)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:

```

; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120638)..(120638)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature

```

```

; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

```

## Alignment Scores:

```

Pred. No.: 7.62e-108      Length: 1830121
Score: 1173.00           Matches: 221
Percent Similarity: 91.00%    Conservative: 42
Best Local Similarity: 76.47%  Mismatches: 26
Query Match: 82.09%          Indels: 0
DB: 16                     Gaps: 0

```

US-09-545-199F-4 (1-289) x US-10-329-670-1 (1-1830121)

```

QY 1 MetAlaGlyAlaLysGluLeuArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
Db 501934 ATGGCAGGTGCAAAAGAGATAAAACCAAAATGCCAGTGTACAGTACACAAAAATC 501875
QY 21 ThrLysAlaMetGluMetValAlaLysLysMetArgLysThrGlnGluArgMetSer 40
Db 501874 ACTAAGGCATGGAATGTGGCAACCTCGAAAAATGCGTAAACGACGATCGTATGGCT 501815
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db 501814 GCATCTCGTCCGTATTCTGAAACTATCCGTAAACGTTATTAGTCATGTGCTTAAGCCAGT 501755
QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
Db 501754 ATCGGTATTAACAATCCGTTCTTAGTTAGCCGCAAGTGAAGAAAAATCGGTATCTTGGTT 501695
QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db 501694 ATTCAACAGATCGTGGGATGTGGTGGTTAATGTTAATTTATTTCAAAACCACTT 501635
QY 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120

```

```

; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29238)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4416)..(4416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a,t,c, or g

```

```

/
/
/ NAME/KEY: misc feature
/ LOCATION: (80024)..(80024)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (100091)..(100091)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (102696)..(102696)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (105121)..(105121)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (107248)..(107248)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (117136)..(117136)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (119750)..(119750)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (119324)..(119324)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (120038)..(120038)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (122167)..(122167)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (122336)..(122336)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (131340)..(131340)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (131360)..(131360)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (139910)..(139910)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (140398)..(140398)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (142750)..(142750)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:

```

```

/
/
/ NAME/KEY: misc feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/
/ NAME/KEY: misc feature
/ LOCATION: (152530)..(152530)

```

## Alignment Scores:

```

Pred. No.:      7.62e-108      Length:      1830121
Score:          1173.00        Matches:      221
Percent Similarity: 91.00%      Conservative: 42
Best Local Similarity: 76.47%    Mismatches:   26
Query Match:      82.09%        Indels:       0
DB:               18           Gaps:          0

```

US-09-545-199F-4 (1-289) x US-10-158-865-1 (1-1830121)

```

Qy      1  MetalaglyalalysGluileArgThrLyleAlaserVallysserThrGlnlySile 20
Db      501934  ATGCAGGTGCAAAAGAGATAAAACCAAAATGCCAGGTACAAAGTACACAAAATC 501875

Qy      21  ThrLysalaMetGluMetValAlaAlaserLysMetArgLysThrGlnGluArgMetSer 40
Db      501874  ACTAAGGCATGGAAATGGTGGCACTCGAAATCGTAAACCGCAGGATCGTATGGCT 501815

Qy      41  SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60
Db      501814  GCATCTCGTCGTATCTGAAACTATCCGTAAAGTTATTAGTCATGTCCTAAGGCAAGT 501755

Qy      61  IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal 80
Db      501754  ATCGTTATAACATCCGTTCTTAGTTGAGCGCGAAGTGAAGAAATCGGTATCTTGGTT 501695

Qy      81  ValSerThrAspArgGlyLeuGlyCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db      501694  ATTCAACAGCATCGTGGGATGTGTGGTGTTAATGTATTATTATTCAAAACACACATT 501635

Qy      101  AsnGluMetLysGluThrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db      501634  AACCAATATAAAATTTGGAAGACANAATATTTCTACAGATTTGGGCTTAATAGTTCA 501575

Qy      121  LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
Db      501574  AAAGGATTAGTTTTTTCGGTTCTTTGGATTATATCAAGGTACGCTTCTGGTTTA 501515

Qy      141  GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db      501514  GCGGATACGCCGCTCTAGAAGAATTAATTTGGTGGCAATACATAATGTTTGAATGCTTAT 501455

Qy      161  LysLysGlyGluValAspValValTyrLysValTyrAsnLysPheIleAsnThrMetSer 180
Db      501454  CGTAATGGTGAATGATGCAATTTATTCATPACAAATAATTTGTTAATACGATGTCG 501395

Qy      181  GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnGluLeuGly 200
Db      501394  CAAAGCCTGTTGTACAACTATTAGTTCCTTTACCAGATCTTAAGACGATCATTTAAAT 501335

```

QY 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAen 220  
 Db 501334 GAAAGCAACAGACTTGGGATTATCTTTATGACCAAGAACCAAAAGTAGTACTATTAGATAGC 501275  
 QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGlnAlaSerLeuAlaSer 240  
 Db 501274 CTTTAGTCTGTTATTAGAGTCCCAATTTATCAAGCGGTTGTAGATAATGTAGCTTCA 501215  
 QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAenAlaGlyAsnLeuIle 260  
 Db 501214 GAACACGCGCTCGAATGTAGCAATGAAAGCAGCAACGATAATGCAGGTAATTTAAT 501155  
 QY 261 AsnGluLeuGlnValTyrLeuLysAlaArgGlnAlaSerIleThrAsnGluLeuAen 280  
 Db 501154 AATGATCTCGGTTGTGTACACAAAGCTGTCAAGCAAGTATCACAAATGAATTAAT 501095  
 QY 281 GluIleValAlaGlyAlaAlaLalle 289  
 Db 501094 GAAATCGTAGCGGTCGCGCGCGATT 501068  
 RESULT 5  
 US-09-809-665A-132  
 ; Sequence 132, Application US/09809665A  
 ; Publication No. US20040110268A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lowery E., David, et al.  
 ; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
 ; FILE REFERENCE: 28341/00435  
 ; CURRENT APPLICATION NUMBER: US/09/809,665A  
 ; CURRENT FILING DATE: 2001-03-15  
 ; PRIOR APPLICATION NUMBER: 60/153,453  
 ; PRIOR FILING DATE: 1999-09-10  
 ; PRIOR APPLICATION NUMBER: 60/128,689  
 ; PRIOR FILING DATE: 1999-04-09  
 ; PRIOR APPLICATION NUMBER: 09/545,199  
 ; NUMBER OF SEQ ID NOS: 197  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 132  
 ; LENGTH: 867  
 ; TYPE: DNA  
 ; ORGANISM: Actinobacillus pleuropneumoniae  
 ; FEATURE:  
 ; OTHER INFORMATION: atpG  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(864)  
 US-09-809-665A-132  
 Alignment Scores:  
 Pred. No.: 2,66e-108 Length: 867  
 Score: 1132.50 Matches: 215  
 Percent Similarity: 89.97% Conservative: 45  
 Best Local Similarity: 74.39% Mismatches: 28  
 Query Match: 79.25% Indels: 1  
 DB: 11 Gaps: 1  
 US-09-545-199F-4 (1-289) x US-09-809-665A-132 (1-867)  
 QY 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
 Db 1 ATGGCAGGTGCGAAGAGATAGAACCAAAATGCAAGTGTGAAATACTCAAAAAATC 60  
 QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 Db 61 ACCAAAGCAATGAAATGTTGTCTACCTCTAAATGCGTAAACCGCAAGCGGTATGGCT 120  
 QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 121 GCCAGTCTGCTTATTCGGAACAATCCGTAAAGGTATAGCCATATTGCGAAAGGAAGC 180  
 QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
 Db 181 ATTGGTTAAGCACCCTGTTTAACTGCAAGCTGATTAATAAAGTAGGCTATCTTGTC 240

QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAenValAenLeuPheLysThrValLeu 100  
 Db 241 GTTTCGACCGATCGCGTTTATGCGTGCCTTAATATCAATTTATTCAAAGCCACTTTG 300  
 QY 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120  
 Db 301 AATGAAATTTAAACCTGGAAGATGAAGACCTTAGTGTGACCTTGGTTAGTAGGCTCG 360  
 QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140  
 Db 361 AAAGCGTAAGCTTTTACCAAAATCTAGGCTTAACGTGAGATCTCAAGTAACGGGATTA 420  
 QY 141 GlyAspThrProSerValGlnGluLeuIleGlySerValAenSerMetIleAspAlaTyr 160  
 Db 421 GCGGATAATCCGGAATGGAACGTATCGTGGCGCAGTTAATGAAATGATTAAATGCGTTC 480  
 QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
 Db 481 CGAAACGGAGAGTGGATGCGGTTTACGTCGCTTACACCGTTTTGAAATACGATGTCA 540  
 QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAenAspGluLeuGly 200  
 Db 541 CAAAAACCTGTTATCGCACAGTTACTTCCGTTACTTAACTAACTAGATGACGATGAATTA 597  
 QY 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAen 220  
 Db 598 GATACGAAAGGTTTCATCGGATTATATTTATGAACCGAATCCACAGTTTATTTGATAGT 657  
 QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaAlaValGluAsnLeuAlaSer 240  
 Db 658 TTACTTCTGTTTATTAGAACTCAGGTATACCAAGCAGTTGTAGATAACCTAGCTTCT 717  
 QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAenAlaGlyAsnLeuIle 260  
 Db 718 GAACAAGCCGCTCGAATGCTAGCGATGAAGCGGCAACAGATAATGCGGTTACATTAATC 777  
 QY 261 AsnGluLeuGlnValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAen 280  
 Db 778 GATGAATTAACATTAATGATGATTAACAAAGCTCCCAAGCAAGCATTACAATGAATTAAC 837  
 QY 281 GluIleValAlaGlyAlaAlaLalle 289  
 Db 838 GAAATGTTGCGGTCGCGCGCGCAATTT 864

RESULT 6  
 US-09-809-665A-166  
 ; Sequence 166, Application US/09809665A  
 ; Publication No. US20040110268A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lowery E., David, et al.  
 ; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
 ; FILE REFERENCE: 28341/00435  
 ; CURRENT APPLICATION NUMBER: US/09/809,665A  
 ; CURRENT FILING DATE: 2001-03-15  
 ; PRIOR APPLICATION NUMBER: 60/153,453  
 ; PRIOR FILING DATE: 1999-09-10  
 ; PRIOR APPLICATION NUMBER: 60/128,689  
 ; PRIOR FILING DATE: 1999-04-09  
 ; PRIOR APPLICATION NUMBER: 09/545,199  
 ; NUMBER OF SEQ ID NOS: 197  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 166  
 ; LENGTH: 866  
 ; TYPE: DNA  
 ; ORGANISM: Pasteurella (Mannheimia) haemolytica  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(864)  
 ; OTHER INFORMATION: atpG  
 US-09-809-665A-166

## Alignment Scores:

Pred. No.: 1.42e-102 Length: 866  
 Score: 1077.50 Matches: 208  
 Percent Similarity: 87.89% Conservative: 46  
 Best Local Similarity: 71.97% Mismatches: 34  
 Query Match: 75.40% Indels: 1  
 DB: 11 Gaps: 1

US-09-545-199f-4 (1-289) x US-09-809-665A-166 (1-866)

```

QY 1 MetAlaGlyAlaLysGluLeuAArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
DB 1 ATGCGAGTCTAAAGAGATAGACCAAAATTCGAAGTTCGTAATACACAAAAATTT 60

QY 21 ThrLysAlaMetGluMetValAlaLaserLysMetArgLysThrGlnGluArgMetSer 40
DB 61 ACCAAGCGATGGAATGTTGCGCATCAAAAATCGTAAACCAAGAGCGTATGCGG 120

QY 41 SerSerArgProTyrSerGluThrIleAArgAsnValIleSerHisValSerLysAlaThr 60
DB 121 GCITTCGCGCTTATGCTGNAAGTATTCGAAGGCATTTAGCCATTTGCGCAAGGTAAAC 180

QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal 80
DB 181 ATTGAGTATAAACACCAATTTTGACCCACGTCGGTAAAAAAGTTGGCTATTAGTA 240

QY 81 ValSerThrAspArgGlyLeuCyGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
DB 241 GTTTCACCGATCGCGGTTTATGTTGGCTTAAATATCAATTTATTTAAACCGTTTAA 300

QY 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLysIleGlySer 120
DB 301 CATGAATTCAAGAAAGAAAGATGACCAAGGTGTTAAGTCTCGACTGCTGTTGGTGGAAAT 360

QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
DB 361 AAAGGATCTCTTTTAAACCAATGGGCTAGAGATTAAAGTTCATATCAATGGAATTG 420

QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
DB 421 GGTGATACACCGCAATGGAAGATTAGTCGTTATGTTATGTTATGTTATGTTATGTTATG 480

QY 161 LysLysGlyGluValAspValValLysValLysValLysValLysValLysValLys 180
DB 481 CCGTGAAGCGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540

QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
DB 541 CAAAAACGACAGTACACAGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 597

QY 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuAspAsn 220
DB 598 GAGCAAACTGTTCTCTGGGATTTATCTCTATGAACCAAAATCCACAGCGTTATTAGACAC 657

QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240
DB 658 TACTGGTTCGTTATTAGATCTCAAGTTATCAGCCAGCTGGTAGATATCTTGGCTCT 717

QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
DB 718 GAACAGGCTGCTGCAATGGTGGCAATGAAGAGCAACCGATTAACCGAGTAACTCTGAT 777

QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280
DB 778 AATGAGTTACAGTTAGTGTATACAAAGCTCGTCAAGCAAGTATTACGAATGAATTAAT 837

QY 281 GluIleValAlaGlyAlaAlaIle 289
DB 838 GAAATTCGCGGTCGCGCAGCAAT 864

```

## RESULT 7

US-10-369-4931

; Sequence 44931, Application US/10369493

; Publication No. US20030233675A1

## ; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10 (52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 44931  
 ; LENGTH: 861  
 ; TYPE: DNA  
 ; ORGANISM: Xenorhabdus nematophilus

US-10-369-493-44931

## Alignment Scores:

Pred. No.: 3.09e-95 Length: 861  
 Score: 1007.00 Matches: 195  
 Percent Similarity: 83.74% Conservative: 47  
 Best Local Similarity: 67.47% Mismatches: 45  
 Query Match: 70.47% Indels: 2  
 DB: 15 Gaps: 1

US-09-545-199f-4 (1-289) x US-10-369-493-44931 (1-861)

```

QY 1 MetAlaGlyAlaLysGluLeuAArgThrLysIleAlaSerValLysSerThrGlnLysIle 20
DB 1 ATGCGCGCGCAAAAGAGATACGTACCAAAATCCCGAGTGTGCAGAACACGCAAAAAATC 60

QY 21 ThrLysAlaMetGluMetValAlaLaserLysMetArgLysThrGlnGluArgMetSer 40
DB 61 ACTAAAGCGATGAGATGCGCGCGTCCAAAATGCTAAACGCGAGGATCGATGCGG 120

QY 41 SerSerArgProTyrSerGluThrIleAArgAsnValIleSerHisValSerLysAlaThr 60
DB 121 GCCAGTCGCTTATGACAGAAACCATTCGCGAGTGTGATTCGACACCTTCGTTAGGTAAT 180

QY 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80
DB 181 CTGGAATACAGACATCCATACCTTCAGCGCGCAAGTTAAGCGTCTGGTACTTGGTT 240

QY 81 ValSerThrAspArgGlyLeuCyGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
DB 241 GTTTCGACAGATGTTGTTTATGTCGGGTTTGAACATTAATTCGTTCAAAAAATTCGTG 300

QY 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLysIleGlySer 120
DB 301 ATAGAAATGAAGACCTGGTCTGATAAAGCGCTTCAAGTTGATTTAGCATTCTGTTGGTGC 360

QY 121 LysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140
DB 361 AAAGCGGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 420

QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
DB 421 GGAGATAACCCATCATTCGCGAGCTGATCGCGCGAGTCAACGTCATCATGATGCAAGCATAT 480

QY 161 LysLysGlyGluValAspValValLysValLysValLysValLysValLysValLys 180
DB 481 GACGAGGCGCTGGGATAATTTGATATAGTGACAAAACCGGTTTCATCAATACATGTCT 540

QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
DB 541 CAGGTTCCGACAAATTCAGTTATTCCTCTGCTGCGCGGAGACGATGAACATTCG--- 597

QY 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuAspAsn 220

```

Db 598 ---AAGAAGAAGTCTCGGATATCTGTATGAACCTGATCCTAAAGCGGTGTGGATACC 654  
 Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240  
 Db 655 CTGCTGCGCGCTATATAGAAATCACAAATTTATACAGGCGTGTGTAACAAACCTGCTAGT 714  
 Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLeu 260  
 Db 715 GAACAGCGCCGACCAATGGTAGCGATGAAGAGCGCGACCGGTAACGGTGGCACTGATC 774  
 Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 775 AAGAGGTTCAGTGTGTTTACAAAGGCTGCTGAGCGCAGCATCCTCAGGAATCACC 834  
 Qy 281 GluLeuValAlaGlyAlaAlaIle 289  
 Db 835 GAAATCGTTTCGGGTGCTTCGGGTT 861  
 RESULT 8  
 US-10-369-493-24123  
 ; Sequence 24123, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 24123  
 ; LENGTH: 861  
 ; TYPE: DNA  
 ; ORGANISM: Xenorhabdus nematophilus  
 US-10-369-493-24123

Alignment Scores:  
 Pred. No.: 1.03e-94 Length: 861  
 Score: 1002.00 Matches: 194  
 Percent Similarity: 83.74% Conservatives: 48  
 Best Local Similarity: 67.13% Mismatches: 45  
 Query Match: 70.12% Indels: 2  
 DB: 15 Gaps: 1

US-09-545-199F-4 (1-289) x US-10-369-493-24123 (1-861)  
 Qy 1 MetAlaGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
 Db 1 ATGCGCGCGCGCAAAAGAGATACGTACCAATCGCAGTGTGCAGAACACGCGAAGAAATC 60  
 Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 Db 61 ACTAAACCGATGAAATGGTGGCGCGGTCCAAATCGTAAACCGCAGACCGCATGGCG 120  
 Qy 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 121 GCCAGCGCTCTTATGCAGATACCATTCGACGGTATGGACACTTGGCTAGGTAGTAAT 180  
 Qy 61 IleGlyTyrLysHisProPheLeuValAspArgGluValLysValGlyMetIleVal 80  
 Db 181 CTGAGTACAAACCACTACCTTGAAGAGCGGTGACGTTAAACGGGTGGGTACGTGTT 240  
 Qy 81 ValSerThrAspArgGlyLeuGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
 Db 241 GTTCTTACTACCGTGTGTTATGTGGCGTTTGACATTAATCTTTCAAAAAATGCTG 300  
 Qy 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerIleGlySer 120

Db 301 ATAGAATGAAGACTGGTCTGATATAAAAGTCCAGGTGGTATTTGGCACTATTGGATCA 360  
 Qy 121 LysSerIleAsnPhePheGlnSerLeuGlyLysIleLysIleLeuThrGlnAspSerGlyIle 140  
 Db 361 AAGCGGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 420  
 Qy 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160  
 Db 421 GGAGATAACCCATCATTCGCCAATTTGATCGGCCGACGTGACGTCATGATGCAAGCATAT 480  
 Qy 161 LysLysGlyGluValAspValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
 Db 481 GACGAGGGCGTCTGGATAAACTGTATAGTACAAACAAGTTCCACAATACATGCTCT 540  
 Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
 Db 541 CAGGTTCCGACAACTCAGTCAGTTATTCCTCTGCGCGGAGACGATGAACACTG--- 597  
 Qy 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220  
 Db 598 ---AAGAAGAAGTCTCGGATATCTGTATGAACCATCTTAAGGCGCTGTGGATACC 654  
 Qy 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240  
 Db 655 TTGCTGCGTCTATATAGAAATCGCAAGTTTATCAGGCGCTGTTGAAAACCTGGCTAGT 714  
 Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLeu 260  
 Db 715 GAACAGCGCGCAGCAATGGTAGCGATGAAGCGCGACTGATTAACGTTGGCACTGATC 774  
 Qy 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 775 AAGAGTTTCAGTGTGTTTACAAAGCTGCTGAGCGCAGCATTAATCTCAGGAATCACC 834  
 Qy 281 GluLeuValAlaGlyAlaAlaIle 289  
 Db 835 GAAATCGTTTCGGGTGCGCTGCGGTT 861

## RESULT 9

US-09-741-669-144  
 ; Sequence 144, Application US/09741669  
 ; Patent No. US20020022718A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Forsyth, R. Allyn  
 ; APPLICANT: Olesen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; TITLE OF INVENTION: Genes identified as required for  
 ; FILE REFERENCE: ELITRA.009A proliferation of E. coli  
 ; CURRENT APPLICATION NUMBER: US/09/741,669  
 ; PRIOR FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/173005  
 ; PRIOR FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 144 -  
 ; LENGTH: 864  
 ; TYPE: DNA  
 ; ORGANISM: Escherichia coli  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(864)  
 US-09-741-669-144

Alignment Scores:  
 Pred. No.: 1.44e-93 Length: 864  
 Score: 991.00 Matches: 191  
 Percent Similarity: 82.35% Conservatives: 47  
 Best Local Similarity: 66.09% Mismatches: 49  
 Query Match: 69.35% Indels: 2  
 DB: 5 Gaps: 1



US-09-545-199f-4 (1-289) x US-09-741-669-144 (1-864)

```
QY 1 MetAlaGlyAlaLysGluLeuArgThrLysLysLeuAlaSerValLysSerThrGlnLysIle 20
Db 1 ATGGCGCGCGCAAAAGAGATAGTAGTAAGATCGCAGCGCTCCAGAACACCGCAAAAGATC 60
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
Db 61 ACTAAGCGATGGAGATGGTGGCGCTCCAAATGCGTAATCGCAGGATCGCATGGCG 120
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValLysSerHisValSerLysAlaThr 60
Db 121 GCCAGCGCTCTTATGCAGAAACCATCGCAAGCTGATGGTACCTTGCCACCGGTAAT 180
QY 61 IleGlyTyrLysHisProPheLeuValAspValSerValGlnLeuSerLeuIleGlySer 120
Db 301 GCGGAATAGAACACCTTACCTGGAAGACCGCGACCTTAAACGCGTGGCTACCTGGTG 360
QY 121 LysSerIleAsnPheGlnSerLeuGlyLysIleLysLysValLysLysValGlyMetIleVal 140
Db 361 AAAGCGGTGCTCTTCACTCCGCGCGGCGCATGTTGTCGCCACCGTCAACCGGATG 420
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 421 GGGGTAACCCCTTCCCTGCGCAACCTGATCGGTAAAGTGTGTTGAAACCTTCAAAA 480
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 481 GACGAAGCGGCTGACCAAGCTTACATTGTTCAGCAAAATTTATTAAACCATGTCT 540
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 541 CAGGTTCCGACCATCAGCGCTGCTGCGCTTACCGGCATCAGATGATGATGATG 240
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 598 ---AAACATAATCTGGGATTACCTGACGAACCGCATCCGAAGCGGCTGCTGATC 654
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240
Db 655 CTGCTCGCTGCTATGTCGAATCTCAGGTTTATCAGGCGCTGTTGAAACCTTGGCCAGC 714
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db 835 CAGATCGTCTCGGGGCGCGCGCGGTT 861
```

## RESULT 10

US-10-369-493-47300  
; Sequence 47300, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 47300  
; LENGTH: 864  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-10-369-493-47300

## Alignment Scores:

Pred. No.: 1,44e-93 Length: 864  
Score: 991.00 Matches: 191  
Percent Similarity: 82.35% Conservative: 47  
Best Local Similarity: 66.09% Mismatches: 49  
Query Match: 69.35% Indels: 2  
DB: 15 Gaps: 1

US-09-545-199f-4 (1-289) x US-10-369-493-47300 (1-864)

```
QY 1 MetAlaGlyAlaLysGluLeuArgThrLysLysLeuAlaSerValLysSerThrGlnLysIle 20
Db 1 ATGGCGCGCGCAAAAGAGATAGTAGTAAGATCGCAGCGCTCCAGAACACCGCAAAAGATC 60
QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40
Db 61 ACTAAGCGATGGAGATGGTGGCGCTCCAAATGCGTAATCGCAGGATCGCATGGCG 120
QY 41 SerSerArgProTyrSerGluThrIleArgAsnValLysSerHisValSerLysAlaThr 60
Db 121 GCCAGCGCTCTTATGCAGAAACCATCGCAAGCTGATGGTACCTTGCCACCGGTAAT 180
QY 61 IleGlyTyrLysHisProPheLeuValAspValSerValGlnLeuSerLeuIleGlySer 80
Db 181 CTGGAATATAGAACACCTTACCTGGAAGACCGCGACCTTAAACGCGTGGCTACCTGGTG 240
QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100
Db 241 GTGTCGACCGCGCTGTTGTCGGTGGTGGTGAACATTAACTGTTCAAAAACCTGCTG 300
QY 101 AsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120
Db 301 GCGGAATAGAACACCTGACCGCGCAAAAGCGTTCATTCGACCTCGCAATGATCGCTCG 360
QY 121 LysSerIleAsnPheGlnSerLeuGlyLysIleLysLysValLysLysValGlyMetIle 140
Db 361 AAAGCGGTGCTCTTCACTCCGCGCGGCGCATGTTGTTGCCCGCATCACCAGGATG 420
QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160
Db 421 GGGGATAACCCCTTCCCTGCGCAACCTGATCGGTAAAGTGTGTTGAAACCTTCAAAA 480
QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180
Db 481 GACGAAGCGGCTGACCAAGCTTACATTGTTCAGCAAAATTTATTAAACCATGTCT 540
QY 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200
Db 541 CAGGTTCCGACCATCAGCGCTGCTGCGCTTACCGGCATCAGATGATGATGATGATG 597
QY 201 GluArgLysGlnValTyrAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220
Db 598 ---AAACATAATCTGGGATTACCTGACGAACCGCATCCGAAGCGGCTGCTGATC 654
QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240
Db 655 CTGCTCGCTGCTATGTCGAATCTCAGGTTTATCAGGCGCTGTTGAAACCTTGGCCAGC 714
QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260
Db 835 CAGATCGTCTCGGGGCGCGCGGTT 861
```

Db 715 GAGCAGCCGCCGCTATGCTGGCGATGAAGAGCCGCGACACATGCGCGAGCTGATT 774  
 Qy 261 AsnGluLeuGlnLeuValTyraenLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 775 AAAGAGCTGCGAGTTGGTATACACAAAGCTCTGAGGCGAGCATTTACTCAGGAACCTACC 834  
 Qy 281 GluIleValAlaGlyAlaAlaAla 289  
 Db 835 GAGATCGTCTCGGGCGCCCGCGGTT 861  
 RESULT 11  
 US-10-369-493-37532  
 ; Sequence 37532, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 37532  
 ; LENGTH: 858  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas fluorescens  
 US-10-369-493-37532  
 Alignment Scores:  
 Pred. No.: 1.45e-84 Length: 858  
 Score: 904.50 Matches: 177  
 Percent Similarity: 78.20% Conservative: 49  
 Best Local Similarity: 61.25% Mismatches: 60  
 Query Match: 63.30% Indels: 3  
 DB: 15 Gaps: 1  
 US-09-545-199F-4 (1-289) x US-10-369-493-37532 (1-858)  
 Qy 1 MetalGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
 Db 1 ATGGCAGCGCCAAAGAGATTCGAGTAAGATTGCGAGCATCAAAGCAGCAGAAAATT 60  
 Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnLysArgMetSer 40  
 Db 61 ACCAGCGCCATGAAAGAGTGGCGGTACAGAAATGCGCAGGCAAAATGCGCATGCGT 120  
 Qy 41 SerSerArgProTyraenLysGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 121 GCTAGCGCGCTTATCGGAGCGGTATCCGCCAGGTAATGGGCATCTCGCCAAACGCCAAC 180  
 Qy 61 IleGlyTyraenLysHisProPheLeuValAspArgGluValLysLysValGlyMetIleVal 80  
 Db 181 CCGGAATACCGCCACCGCTTCATGATCGATCCGGAAGTAAAGCGCGTGGTATGCGTA 240  
 Qy 81 ValSerThrAspArgGlyLeuGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
 Db 241 GTGAGCAGTGACCGTGGTGTGTGCGCGCGCTTGACACCAACCTGTTCAAGCGCCCTGTC 300  
 Qy 101 AsnGluMetLysGluTyraenLysLysAspValSerValGlnLeuSerLeuIleGlySer 120  
 Db 301 AAGGACATGCGCGTAAACCGCGAAACCGCGTCGAGATCGATCTGTGCTGTGCTGAGC 360  
 Qy 121 LysSerIleAsnPhaPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGlyIle 140  
 Db 361 AAGGTGCGCGCTTTTCCGAACTTCGCGCGTAACTGCTGAGTGTGCGATCATCAGCCACTG 420  
 Qy 141 GlyAspThrProSerValGlnLeuIleGlySerValAsnSerMetIleAspAlaTyra 160

Db 421 GGTGAAGAGCCGCTGATCAATGATCTGATCGCAGCGTCAAGGTGATGCTGGATGCTAT 480  
 Qy 161 LysLysGlyGluValAspValValTyraenLysLysValTyraenLysLysLysLysLys 180  
 Db 481 CTGGACGCGCGTATGACCGCTGTCTGCTGTATCCCAACAAAGTTTCATCAACACCATGACG 540  
 Qy 181 GlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeuGly 200  
 Db 541 CAACAGCCTACCGTGGAGCAGTTGATTCATCGTGGCGACCCCGGATCAAGAACTC--- 597  
 Qy 201 GluArgLysGlnValTyraenLysLysLysLysLysLysLysLysLysLysLysLysLys 220  
 Db 598 -----AAGCACCCTGGGACTACCTCTACGAAACGAGCGCAAGAGCTGCTTGCAGCG 651  
 Qy 221 LeuLeuValArgTyraenLysLysLysLysLysLysLysLysLysLysLysLysLysLys 240  
 Db 652 TTGATGCTGGCTTACGTGGAGTCCGAGGTGTACAGCGCGTGTGCGAGAACACCGCGCT 711  
 Qy 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuIle 260  
 Db 712 GAACAGCTGCGCGGATGATCGGATGAAGACGCTACCGCAACGCGCTGATTTGATC 771  
 Qy 261 AsnGluLeuGlnLeuValTyraenLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 772 AGCGATTTGCGAGCTGATCTACAAAGCGCGTACCGCAACGCGCTGATTTGATC 831  
 Qy 281 GluIleValAlaGlyAlaAlaAla 289  
 Db 832 GAAATCGTGGCGCGCTGCGCGGTT 858  
 RESULT 12  
 US-10-369-493-32530  
 ; Sequence 32530, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 32530  
 ; LENGTH: 867  
 ; TYPE: DNA  
 ; ORGANISM: Ralstonia metallidurans  
 US-10-369-493-32530  
 Alignment Scores:  
 Pred. No.: 3.75e-83 Length: 867  
 Score: 891.00 Matches: 171  
 Percent Similarity: 78.89% Conservative: 57  
 Best Local Similarity: 59.17% Mismatches: 59  
 Query Match: 62.35% Indels: 2  
 DB: 15 Gaps: 1  
 US-09-545-199F-4 (1-289) x US-10-369-493-32530 (1-867)  
 Qy 1 MetalGlyAlaLysGluIleArgThrLysIleAlaSerValLysSerThrGlnLysIle 20  
 Db 1 ATGGCAGCGCCAAAGAGATTCGAGTAAGATTGCGAGCATCAAAGCAGCAGAAAATT 60  
 Qy 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnLysArgMetSer 40  
 Db 61 ACCAAGCGCGATGAGATGTCGCGCGATCCAGATCGCGCAGGCGCAGGACGATCGC 120

QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerIysAlaThr 60  
 DB 121 AACGCCGCCCTACGGCGAAGTCCGGAATATCCGGCCACCTGGCCCTCGCCAC 180  
 QY 61 IleGlyTyrIysHisProPheLeuValAspArgGluValIysLysValGlyMetIleVal 80  
 DB 181 CCCGAGTTCAAGCATCGGTTTCATGGTGGCACCGGACGTCAGAGCGTCGGCATGTCGTG 240  
 QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheIysThrValLeu 100  
 DB 241 GTGACGACCCACCAAGGTCGTGCGGTGGCTGACCAACCAACGTCGTGCGCGGTGACC 300  
 QY 101 AsnGluMetIysGluTrpIysGluIleValSerValGlnLeuSerIleGlySer 120  
 DB 301 AACCAACTGAAGACCTGCAGCGGCCAGCGGCTGAACGTGCAAGCGACCGCCATCGGTACC 360  
 QY 121 LysSerIleAsnPhePheGlnSerLeuGlyIleLeuIleLeuThrGlnAspSerGlyIle 140  
 DB 361 AAGGGCATGAGTTCCTGGCCCATCGCGCCGCAAGGTGGTCTCGCATGTGGTGCAGGTG 420  
 QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaTyr 160  
 DB 421 GGTGACACCCCGCATCTGGAAGCTGATCGCGCGCATCAAGTTTCACCTGCGACGCTAT 480  
 QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
 DB 481 ACCAACGGGAAGTGCACCGGTGACTTGGCATATACCAAGTTTCATCAACACGATGAAG 540  
 QY 181 GlnLysProValLeuGluLysLeuIleProLeu-----ProGluLeuAspAsnAspGlu 198  
 DB 541 CAGGAGCCGATGTCGAGCACTGTCGCGCTGGCAGCGGCAAGCTGAGCCAGCCGAA 600  
 QY 199 LeuGlyGluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeu 218  
 DB 601 GATGAAGAGCGGCTACTCGTGGGATTTACATCTACGAGCCGCGACGCGCCAGCGTTGTG 660  
 QY 219 AspAsnLeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaLysValGluAsnLeu 238  
 DB 661 GAAGAGTGTCTGTCGCTAGCTCGAAGCGTGTGTACAGCGCGTGGCGGAGACATG 720  
 QY 239 AlaSerGluGlnAlaAlaArgMetValAlaMetIysAlaAlaValAsnAlaGlyAsn 258  
 DB 721 CGCTCGGAGCAATCGCGCGTATGTCGCCATGAAGCTGCATCGACACACCGCAAGAAC 780  
 QY 259 LeuIleAsnGluLeuGlnValTyrAsnLysAlaAlaArgGlnAlaSerIleThrAsnGlu 278  
 DB 781 GTGATCGCGCACTGCACACTGCTCTACAAAGACCCGTCAGGACGATACGAAGAA 840  
 QY 279 LeuAsnGluIleValAlaGlyAlaAla 287  
 DB 841 CTGTGGAAATCGTCAGCGGTGCGCT 867

## RESULT 13

US-10-369-43286  
 ; Sequence 43286, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR FILING DATE: 2002-02-21  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 43286  
 ; LENGTH: 862  
 ; TYPE: DNA

; ORGANISM: Nitrosomonas europaea  
 US-10-369-493-43286

## Alignment Scores:

Pred. No.: 1,44e-82 Length: 882  
 Score: 885.50 Matches: 165  
 Percent Similarity: 78.91% Conservative: 67  
 Best Local Similarity: 56.12% Mismatches: 57  
 Query Match: 61.97% Indels: 5  
 DB: 15 Gaps: 2

US-09-545-199F-4 (1-289) x US-10-369-493-43286 (1-882)

QY 1 MetAlaGlyAlaLysGluIleArgThrIysIleAlaSerValIysSerThrGlnIysIle 20  
 DB 1 ATGCCAGCAGCAGAGATAGCAATAAATCAAAACGGTTAAAATACCAAAAGATT 60  
 QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 DB 61 ACAGTGGCCATGGAATGGTTGCCCTTCCAAATAGAAAGCGCAGGATCGTATGAAG 120  
 QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 DB 121 AAAGCGGTCCTTATGGCGAGAGATACCAACGTGGCTGCACATATGAGTAATGCCAGT 180  
 QY 61 IleGlyTyrLysHisProPheLeuValAspArgGlu---ValLysLysValGlyMetIle 79  
 DB 181 GTTGTAGTATCCCATCCCTCTTCTGATAAGTCTGACTCTGTAAAGAGTGGGATCATTT 240  
 QY 80 ValValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrVal 99  
 DB 241 GTTGTAACTTCAGATAAAGTTTGTGTGGCGGTGAAACACAGTAATGTGTCGTAGAGCA 300  
 QY 100 LeuAsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGly 119  
 DB 301 CTCAATGAATTCGAAGCTGGAAACCGAAGGTAATCATGTAGATGCTGCTGCATAGGT 360  
 QY 120 SerLysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGly 139  
 DB 361 AACAGGGTTGGGTTTCATAGCCGCTTGGTACTCAGGTAATCTCACAAGTAACCGGG 420  
 QY 140 IleGlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAla 159  
 DB 421 TTGGTGTATGCATCACTAATATGGAAGCTTGAATGGTGCAGTCAAGTAGTCTTGGATGCA 480  
 QY 160 TyrLysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMet 179  
 DB 481 TATACAGAGCGCAGTTGGATAGGTTATATTTCTATAACCGCTTTATTATAACAATG 540  
 QY 180 SerGlnLysProValLeuGluLysLeuIleProLeuProGlu-----LeuAsp 195  
 DB 541 AAGCAGATGCGGTTATGGAGCAGCTTCTCCCTTGATGATGATCGTATCAGTAGTAG 600  
 QY 196 AsnAspGluLeuGlyGluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLys 215  
 DB 601 GATGGGAAGCCAGACCTACCCGTCAGCGTGGATTATATTTATGAACCTGAGCAAAA 660  
 QY 216 ValLeuLeuAspAsnLeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaVal 235  
 DB 661 CCGTTATCGCAGATATCATGTCAGGTATATAGAGCGCTCGTATACCGGCAAGTGTCT 720  
 QY 236 GluAsnLeuAlaSerGluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsn 255  
 DB 721 GAGACATGCTTCAGAAACAATCGGCAAGGATGGTTGCAATGAAAGCGGCACTCTGATAAT 780  
 QY 256 AlaGlyAsnLeuIleAsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIle 275  
 DB 781 GCAGGTAACTGATTTGATGAGCTGACGCTCAITTTATACCAAGTCTCTGTCAGGACCAT 840  
 QY 276 ThrAsnGluLeuAsnGluIleValAlaGlyAlaAlaAla 289  
 DB 841 ACCAAGGAATTTGCAGAAATCGTGGCGGTGCTGCTGCTGTT 882

## RESULT 14

US-10-672-787-38  
 ; Sequence 38, Application US/10672787  
 ; Publication No. US20040067554A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LAGACE, Robert, E.  
 ; APPLICANT: PATTERSON, Chandra  
 ; APPLICANT: BERG, Kim, L.  
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME  
 ; FILE REFERENCE: EPITRA.025C1  
 ; CURRENT APPLICATION NUMBER: US/10/672,787  
 ; CURRENT FILING DATE: 2003-09-26  
 ; PRIOR APPLICATION NUMBER: 09/596,002  
 ; PRIOR FILING DATE: 2000-06-16  
 ; NUMBER OF SEQ ID NOS: 41  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 38  
 ; LENGTH: 94750  
 ; TYPE: DNA  
 ; ORGANISM: Moraxella catarrhalis  
 US-10-672-787-38

## Alignment Scores:

Pred. No.: 1,16e-79 Length: 94750  
 Score: 885.00 Matches: 164  
 Percent Similarity: 78.55% Conservative: 63  
 Best Local Similarity: 56.75% Mismatches: 62  
 Query Match: 61.93% Indels: 0  
 DB: 16 Gaps: 0

US-09-545-199F-4 (1-289) x US-10-672-787-38 (1-94750)

QY 1 MetAlaGlyAlaLeuGluLeuArgThrLysLeuAlaSerValLysSerThrGlnLysLeu 20  
 Db 74372 ATGGCAAGCTTAAAGAAATTCGTCTAAAGTACCGAGCTTAAAGTACCGAAAGAT 74431  
 QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 Db 74432 ACGCGTGAATGCGATGCGTGGCTGCCAGTAAGATGCGTCCCAAGAGAGATGGAA 74491  
 QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 74492 TTGGTCTGCTCTTATTCAGATGGTATACCGCGTGCATTCACATTTGGTGAAGCGCAA 74551  
 QY 61 IleGlyTyrLysHisProPheLeuValAspValSerValGlnLeuSerLeuIleGlySer 80  
 Db 74552 TCAGATTATAGCATCCCTATATGATTAACCGTCCGTCATCGTCTTGGTTTGGTT 74611  
 QY 81 ValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrValLeu 100  
 Db 74612 GTAACCTCTGACCGTGGTTTGGCAGGTGGTTTGACATCAATCTATTCAAAACCTGTGA 74671  
 QY 101 AsnGluMetLysGluTrpLysGluLysAspValSerValGlnLeuSerLeuIleGlySer 120  
 Db 74672 AAAACAGTCAAAAGCTATCAAGACGAGTCAAGTGAATTAATTCAGTAAATAGTGCA 74731  
 QY 121 LysSerIleAsnPhePheGlnSerLeuGlyIleLysLeuThrGlnAspSerGlyIle 140  
 Db 74732 AAAGCGTGGCGCTTTTAAAGACTTTGGCGTAGAGTAACCTCAGCTGTCACTACTAT 74791  
 QY 141 GlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAlaThr 160  
 Db 74792 GCGGTAACCCCTGCATTCAGCAATTAATACACCATTCACCATCTGTGATGATAT 74851  
 QY 161 LysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMetSer 180  
 Db 74852 CTCATAGGAATAGATCGCATTTACTTGGTGTATACCATGTTTATCAATGCTATGGCT 74911  
 QY 181 GlnLysProValLeuGlnLysLeuIleProLeuProGluLeuAsnAspGluLeuGly 200  
 Db 74912 CAAATAACCGGTGGTGAACAGATTGTACCATTTGGCAGAGAGTGTAGTTGATGACACTGAG 74971  
 QY 201 GluArgLysGlnValTrpAspTyrIleTyrGluProAspAlaLysValLeuLeuAspAsn 220

Db 74972 CTTGAGGCACACAGTTGGGATTATATTATGAGCCGACACCAAAACCTCATTTGATAGC 75031  
 QY 221 LeuLeuValArgTyrLeuGluSerGlnValTyrGlnAlaValGluAsnLeuAlaSer 240  
 Db 75032 CTACTATTGCGTTACATTGATCTGTAGTCTACCAATCCGTCGAGAGAAATTGCGCTCT 75091  
 QY 241 GluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAlaGlyAsnLeuLeu 260  
 Db 75092 GAACAGTCAGCCCGTATGCTGGCGATGAACAGCGACAGATAATGCTGGTAATTAAT 75151  
 QY 261 AsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThrAsnGluLeuAsn 280  
 Db 75152 AAAGATTACAAATGCTTTATAAAGCTCAGACAGCGAGCGATTACACGAGAAATCTCG 75211  
 QY 281 GluIleValAlaGlyAlaAlaAlaIle 289  
 Db 75212 GAAATCGTTGGTGGTGGCTGGCTGTT 75238

## RESULT 15

US-10-369-493-39109  
 ; Sequence 39109, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 39109  
 ; LENGTH: 861  
 ; TYPE: DNA  
 ; ORGANISM: Xanthomonas campestris  
 US-10-369-493-39109

Alignment Scores:  
 Pred. No.: 1.23e-77 Length: 861  
 Score: 838.00 Matches: 158  
 Percent Similarity: 77.47% Conservative: 69  
 Best Local Similarity: 53.92% Mismatches: 56  
 Query Match: 58.64% Indels: 10  
 DB: 15 Gaps: 3

US-09-545-199F-4 (1-289) x US-10-369-493-39109 (1-861)

QY 1 MetAlaGlyAlaLeuGluLeuArgThrLysLeuAlaSerValLysSerThrGlnLysLeu 20  
 Db 1 ATGGCAGCGCGGAGCGAATCAACACCAAGATCAAGCGTGCAGAACACCCGCAAGTG 60  
 QY 21 ThrLysAlaMetGluMetValAlaAlaSerLysMetArgLysThrGlnGluArgMetSer 40  
 Db 61 ACGCGCGCGCTCGAATATGCTCGCGCTCCCAAGATCCGCAAGCGCAGGACCGCATGAAG 120  
 QY 41 SerSerArgProTyrSerGluThrIleArgAsnValIleSerHisValSerLysAlaThr 60  
 Db 121 ACCTCGCTCTCTACGCGCAGCGCATTAACAGCGTGTATCGGGCATCTGGCGCAGGCGAGC 180  
 QY 61 IleGlyTyrLysHisProPheLeuValAspValGlu---ValLysLysValGlyMetIle 79  
 Db 181 ACCGATACACGATCCGCTTCTGTTGGAGCGTGAAGCGTCAAGCGGTTCGCTACATC 240  
 QY 80 ValValSerThrAspArgGlyLeuCysGlyGlyLeuAsnValAsnLeuPheLysThrVal 99  
 Db 241 GTGATCTCTCCGACCCCGCGCTGGCGCGGCTCTGAACAAACACCTGTTCGCAAGATG 300

```
QY 100 LeuAsnGluMetLysGluTyrLysGluLysAspValSerValGlnLeuSerLeuIleGly 119
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 CTGGCGAAGTGGCCCGTGGCAGCAGACGAGTCCGAGATCGACGTGGTCACCATCGGT 360
QY 120 SerLysSerIleAsnPheGlnSerLeuGlyIleLysIleLeuThrGlnAspSerGly 139
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 CAGAAGGCTCGGGCTTCTTCGGCCGATCAAGTCAACATGTCGGCAGCGTCACGCAT 420
QY 140 IleGlyAspThrProSerValGluGlnLeuIleGlySerValAsnSerMetIleAspAla 159
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 CTGGCCACAGCCCGCACATCGAGCAGTTGGTGGCGTGTCAAGGTGATGCTGGATGCC 480
QY 160 TyrLysLysGlyGluValAspValValTyrLeuValTyrAsnLysPheIleAsnThrMet 179
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 TTCACCGAGGCAAGTGGACCGGTGATCTGCTGTACACCGCTTGTGAACACCATG 540
QY 180 SerGlnLysProValLeuGluLysLeuIleProLeuProGluLeuAspAsnAspGluLeu 199
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 ACGCAGAGCCAGCTTCGAGCAGTTGCTGCCGCTTCGG-----GCC 582
QY 200 GlyGluArgLysGlnVal-----TrpAspTyrIleTyrGluProAspAlaLysVal 216
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 GCCGAGCACAAGTGGCGCACCGACTGGGACTACCTGTACGAAACCGATGCCGCCACC 642
QY 217 LeuLeuAspAsnLeuValArgTyrLeuGlnSerGlnValTyrGlnAlaAlaValGlu 236
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
643 GTCTGGAGCACGTGATGACGGCTACATCGAGTGGTGTACCGAGCGGTGTGGAA 702
QY 237 AsnLeuAlaSerGluGlnAlaAlaArgMetValAlaMetLysAlaAlaThrAspAsnAla 256
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
703 AACGTGGCTCCGAGCAGCGCGCGCATGTGGCCATGAGGGCGCCGAGCAACGCC 762
QY 257 GlyAsnLeuIleAsnGluLeuGlnLeuValTyrAsnLysAlaArgGlnAlaSerIleThr 276
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 AACAGATGATCGGCACCTTGCACTGTCTACAAAGCGCGCCAGGGCGGATCACC 822
QY 277 AsnGluLeuAsnGluIleValAlaGlyAlaAlaIle 289
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
823 CAGGAAATTCGGAATCGTAGCGCGCGCGCGCGCGTA 861
```

Search completed: November 9, 2004, 22:10:43  
Job time : 2233 secs